ifo WORKING PAPERS

402
2023
September 2023

Topic Salience and Political Polarization: Evidence from the German "PISA shock"

Pietro Sancassani

ifo INSTITUTE
Leibniz Institute for Economic Research
at the University of Munich

Imprint:

ifo Working Papers
Publisher and distributor: ifo Institute – Leibniz Institute for Economic Research at the University of Munich
Poschingerstr. 5, 81679 Munich, Germany
Telephone + 49(0)89 9224 0, Telefax +49(0)89 985369, email ifo@ifo.de
www.ifo.de

An electronic version of the paper may be downloaded from the ifo website
www.ifo.de

ifo Working Paper No. 402

Topic Salience and Political Polarization: Evidence from the German “PISA shock”*

Abstract

Does the salience of a topic affect polarization in related parliamentary debates? When discussing a salient topic, politicians might adopt more extreme stances to gain electoral consensus. Alternatively, they could converge towards more moderate positions to find a compromise. Using parliamentary debates from the 16 German state parliaments, I exploit the exogenous increase in the salience of education induced by the unexpectedly low performance of German students in the PISA 2000 test—the German “PISA shock”. I combine machine-learning and text analysis techniques to obtain topic-specific measures of polarization of parliamentary debates. In a difference-in-differences framework, I find that the PISA shock caused an 8.8% of a standard deviation increase in polarization of education debates compared to other topics. The effect is long-lasting and fades after about six years.

JEL: D72, D71

Keywords: polarization, text analysis, machine learning, Germany, PISA shock

Pietro Sancassani
ifo Institute – Leibniz Institute for
Economic Research at the University of Munich
Poschingerstr. 5
81679 Munich, Germany
sancassani@ifo.de

* For valuable comments and discussion, I would like to thank Raphael Brade, Davide Cantoni, Ruben Durante, Jon Fiva, Lukas Mergele, Massimo Morelli, Niklas Potrafke, Ludger Woessmann, Ulf Zoelitz and seminar participants at the CESifo/ifo Junior Workshop on Big Data, 4th MWZ Text-as-Data Workshop, IIPF, IWAEE, EffEE workshop, IAB Workshop and OCCAM meetings. I thank Marc Piopiunik for his contribution to this project. For excellent research assistance, I thank Melchor De La Cruz, Silas Kaatz, and Vera Freundl.

1. Introduction

Little is known about the relationship between the salience of a topic and the polarization in related political debates. Understanding this relationship is crucial since the salience of a topic, namely the amount of attention that it receives, can be manipulated. Traditional and digital media, for example, are prone to presenting reported events in a sensationalized way (Ryu 1982; Soroka et al. 2018; Bleich and van der Veen 2021; Kayser and Peress 2021; Berger 2022). Social media can exacerbate this phenomenon through the "echo chambers" they tend to create (Sunstein 2018; Settle 2018), thus contributing to an increase in the perceived salience of various issues. At the same time, there is mounting evidence of a historically high ideological divide observed in the United States (Bonica 2013; McCarty, Poole, and Rosenthal 2016; Gentzkow, Shapiro, and Taddy 2019; Iyengar et al. 2019) as well as in other countries (Boxell, Gentzkow, and Shapiro 2022). Particularly in parliamentary debates, the phenomenon of polarization has received considerable attention in recent years (Peterson and Spirling 2018; Gentzkow, Shapiro, and Taddy 2019; Goet 2019; Salla 2020; Fiva, Nedregård, and Øien Henning 2022; Lewandowsky et al. 2022). This literature has mostly provided descriptive evidence on its evolution in different countries, but it has been surprisingly silent on *why* it occurs. The aim of this paper is to provide causal evidence on the salience of a topic as a potential determinant of polarization in related parliamentary debates.

Theoretically, it is an open question in which direction topic salience might affect polarization of parliamentary debates. If the salience of a topic increases, parties might pursue a median voter strategy to appeal to more centrist voters, thus resulting in less polarized debates. The theoretical foundation for this argument follows Downs' (1957) seminal work on the *median voter theorem*. Conversely, parties might exploit the increased salience of a topic to amplify their ideological distinctiveness, which would lead to an increase in polarization. Such behavior would be consistent with the *cleavage theory* framework, which dates back to Lipset and Rokkan (1967). Empirically, it is hard to establish whether topic salience affects polarization as, for example, politicians are known to focus on divisive issues (Ash, Morelli, and van Weelden 2017), which would lead to reverse causation.

To test whether topic salience affects the polarization of parliamentary debates, I leverage a natural experiment that led to an increase in the salience of a specific topic: education. I exploit the release of the results of the first *Programme for International*

Assessment (PISA) in December 2001 in the context of German state parliaments. Due to the unexpectedly low performance of German students and the media attention that this event received, this event was soon renamed the "PISA shock". I focus on the parliamentary debates of all German state parliaments for the period 2000-2008, which I have collected and digitized for this project. These debates constitute a novel data source, and the German context provides an ideal setting for my analysis. Germany is a federal country, where each of its sixteen states has its own parliament with exclusive legislative authority on a set of topics, including education. Hence, state-level parliamentary debates about education are policy-relevant and abundant in this context.

Empirically, I combine machine-learning algorithms and text analysis techniques to classify the topic of each speech in the parliamentary debates and compute topic-specific polarization measures. I use a supervised machine-learning model to classify speeches about the main topic of interest: education. I then classify the topics of all the other speeches with an unsupervised machine-learning algorithm, the correlated topic modelling (CTM) (Blei and Lafferty 2007). Using a measure of text similarity, the cosine similarity, I compute topic-specific measures of polarization, which is defined as the extent to which opinions on an issue are opposed across parties. Assuming that expressing different opinions requires people to use different words, more polarized speeches will be less similar. My main measure of polarization is therefore the dissimilarity between speeches from a benchmark party and speeches from other parties on the same topic.

Identifying the impact of salience on the polarization of education debate is challenging because polarization evolves over time. I therefore conduct a difference-in-differences analysis, where the debates on topics other than education act as the counterfactual group. This approach enables me to control for fluctuations in the general level of polarization in parliamentary debates due to time trends or other unrelated factors, such as upcoming elections or the idiosyncratic compositions of the parliaments. I find that topic salience induced by the PISA shock had a substantial impact on parliamentary debates. First, I find a 22% increase in the share of speeches about education following the PISA shock. Second, I find a sizable increase in polarization of parliamentary debates about education equivalent to 8.8% of a standard deviation (SD). The impact corresponds to about 18% of the average polarization between the main center-right (CDU/CSU) and center-left (SPD) parties in the German political landscape. Using an event-study specification, I show that the shock also had a long-lasting impact. It took roughly six years for polarization in education debates to go back to its pre-shock level.

The interaction between members of parliament's (MPs) party affiliation and the treatment status reveals that the increase in polarization is driven by a cleavage between the main center-right (CDU/CSU) and center-left (SPD) parties. Overall, this result aligns well with a *cleavage-theory* framework, where the main parties drift away from each other in their rhetoric over a subject matter.

While the salience of education undoubtedly increased because of the PISA shock in Germany, it is also possible that the increase in polarization was driven by the information revealed by the release of the PISA results. I address this issue by exploiting an additional feature of this setting: the release of state-specific PISA results in June 2002. This event showed large heterogeneities in performance across German states, with the best performing states in Germany placing themselves among the top performing countries. Nonetheless, I do not find significant heterogeneities in the impact of the shock on polarization with respect to the performance of each state. Further, state-specific results were not released for two states, Berlin and Hamburg, and I also do not find any heterogeneities for these states. These findings seemingly suggest that the salience of the topic, rather than the actual performance of the students, affected the polarization of parliamentary debates. Further, I find that the PISA shock also had a positive impact on the number of proposed bills about education, and the impact is driven by rejected bills.

I also provide suggestive evidence on the issues that likely caused an increase in the polarization about education debates. I develop a polarization score to capture terms that are disproportionally used by MPs of one party. Terms that refer to prominent issues at the time of PISA shock, such as developing a monitoring system of student achievement, "all-day schools", and the tracking system, feature among the most polarized terms. This suggests that debates about such issues contributed to the increase in polarization in education.

This study contributes to two strands of the literature. First, I contribute to the growing literature investigating political polarization. Most studies in this field have focused on the determinants of polarization among voters. This strand of research has shown a relationship between the rise in political polarization and rising import competition (Autor et al. 2020), intensified media partisanship (DellaVigna and Kaplan 2007; Levendusky 2013; Prior 2013), and financial crises (Mian, Sufi, and Trebbi 2014;

Funke, Schularick, and Trebesch 2016). A polarized electorate can lead to more polarization in parliamentary debates, but this link is far from being established in the literature. In fact, causal evidence on the determinants of polarization in the context of parliamentary debates is largely absent. This is surprising given the outburst of studies documenting polarization in parliamentary debates observed in the last years, with evidence from the US (Jensen et al. 2012; Lauderdale and Herzog 2016; Gentzkow, Shapiro, and Taddy 2019), the UK (Peterson and Spirling 2018; Goet 2019), Germany (Lewandowsky et al. 2022), Norway (Fiva, Nedregård, and Øien Henning 2022), and Finland (Salla 2020).[1] I therefore contribute to this literature by providing causal evidence of the effect of topic salience on polarization in parliamentary debates.

Second, I contribute to the political economy of education literature. I show that international standardized assessments, such as PISA, can influence the political discourse about education. Other studies have highlighted the role of interest groups, unions (McDonnell and Weatherford 2013; Galey-Horn et al. 2020), and teacher strikes (Lyon and Kraft 2021) in shaping education policymaking. Public opinion and interest groups are often considered to have a greater role in shaping education policy than insights drawn from empirical data (West and Woessmann 2021). I challenge this notion by providing evidence on the far-reaching consequences of the introduction of an international standardized assessment on the policy-making debate about education. A likely reason behind the impact of the PISA shock is that PISA introduced accountability for policymakers in education. Accountability has been often cited as a key factor to improve the quality of education systems (Woessmann 2009; Figlio and Loeb 2011; Global Education Monitoring Report Team 2017; Bergbauer, Hanushek, and Woessmann 2021). In fact, the lack of comparable student assessments in many countries prevented policymakers from being held accountable for students' performance. This dramatically changed after PISA, as the strong reaction of German policy makers clearly illustrates. The influence of PISA, and the PISA shock, for policymaking in education in various countries has been widely acknowledged in the literature.[2] To the best of my knowledge, no study has attempted to establish a causal

[1] Using US congressional vote choices rather than parliamentary debates, Canen, Kendall, and Trebbi (2020a, 2020b) highlight the role of party discipline as a driver of political polarization.

[2] A vast literature has discussed the implication of PISA for education policy in various countries (Rinne, Kallo, and Hokka 2004; Grek 2009; Bieber and Martens 2011; Breakspear 2012; Martens and Niemann 2013, among others). Several studies have also investigated the consequences of the PISA shock in Germany (Tillmann 2004; Ertl 2006; Waldow 2009; Neumann, Fischer, and Kauertz 2010; Davoli and Entorf 2018, among others).

relationship between PISA results and the political debate about education. I therefore fill the gap in this literature by providing causal evidence on how the international standardized assessment can shape education policymaking.

The remainder of this paper is structured as follows. In Section 2, I provide details about the PISA shock, the concept of topic salience, and the German political system. In Section 3, I present the data and methods used to compute the polarization measures as well as descriptive statistics. In section 4, I present the empirical strategy. In section 5, I report the main results and robustness checks. I provide evidence on the polarizing issues in Section 6. Section 7 concludes.

2. Institutional Background

2.1. The PISA shock

The publication of the results of the first PISA study on the 4th of December 2001 was a watershed in the discourse on education in Germany. The poor and largely unequal performance of German students in PISA sparked heated public debates, with newspaper headlines such as "Catastrophic Results for German Students" (FAZ 2001), "A Disaster in Almost Every Respect" (TAZ 2001), or "Are German students stupid?" (Der Spiegel 2001) populating German newspapers for months. In the two months after the publication of the PISA results, the OECD calculated that daily and weekly newspapers published 774 pages of printed article about this event in Germany, compared to 8 in Finland, the "PISA champion country", 32 in France, whose placement was well above Germany in the PISA ranking, and 16 in Italy, whose performance was akin to Germany (Hopmann, Brinek, and Retzl 2007). The "tsunami-like" impact of this event in Germany (Gruber 2006) was so great that it was soon dubbed the PISA shock and its consequences shaped the public and political debate about education in the following years. In June 2002, roughly six months after the PISA shock, results for German federal states were published and revealed large differences in achievement between the states.[3] Although there were already some indications of such heterogeneities (Ebenrett, Hansen, and Puzicha 2003), this event further fueled the already heated debate about education.

Several reasons lie behind the stir caused by the publication of the first PISA results. First, PISA contradicted the public's perception of the German education system, an

[3] Results were published for all states but Berlin and Hamburg, which did not meet the required criteria for overall reporting (Artelt et al. 2002). State-specific results are reported in Table A1.

assessment that was characterized by self-confidence and belief in its efficiency, which reflected the strong country's economy (Sloane and Dilger 2005; Davoli and Entorf 2018). Second, it represented a threat to a major exporting economy that relies on human capital and skills for its competitive advantage. Third, PISA, and the *International Mathematics and Science Study* (TIMSS) before it, ended a long phase of German abstention from international large-scale assessments (Waldow 2009). In fact, Germany's participation and low performance in the first TIMSS study in 1995 was the first wake-up call for the German education system, but this event, unlike PISA, was largely ignored by the German media (OECD 2011). Germany's decade-long abstention from international assessments was in line with educators' mainstream paradigm that "what is important about education cannot be measured" (Bos and Postlethwaite 2002). PISA abruptly ended this phase, and Germany committed itself to participating in international assessments for years to come.

The PISA shock provided a formidable impetus for reforms in the German education systems. While an exhaustive exposition of such reforms is outside the scope of this paper,[4] they mostly revolved around three areas: developing a monitoring system with common educational standards and central examination, expanding "all-day school" offers, and reforming the tracking system.

2.2. Topic Salience

In this section, I clarify the concept of salience, which plays a crucial role in my analysis. I adhere to the concept of salience defined in a recent review of the literature that studies the role of salience in economic choice by Bordalo, Gennaioli, and Shleifer (2022). The authors describe salience as "the property of a stimulus that draws attention bottom up" (p. 524). Psychologists differentiate between top-down and bottom-up attention as the two methods through which human minds select what to focus on. Top-down attention is voluntary and is the result of an active cognitive process, whereas bottom-up attention is involuntary and occurs automatically. Bordalo, Gennaioli, and Shleifer (2022) identify three factors that make a stimulus salient: contrast with surroundings (contrasting), surprise, and prominence.

It is easy to reconcile this definition of salience with the PISA shock. First, the PISA shock can be identified as a stimulus that drew public attention toward education bottom up, as it came as a reaction to the information made available by the PISA study.

[4] Interested readers may find detailed accounts of these in Ertl (2006), Gruber (2006), Waldow (2009), OECD (2011), and Davoli and Entorf (2018), among others.

Second, the three factors that make a stimulus salient accurately depict the PISA shock: contrast with surroundings, surprise and prominence. An important feature that emerged from the first PISA results was that Germany was a country below the OECD average in terms of student test scores. This element of comparison with other countries—*contrast with surroundings*—contributed to the prominence that the publication of the first PISA results received. As argued in the previous section, PISA revealed a picture of the German education system that was largely unexpected and, therefore, *surprising*. Further, the PISA shock was very *prominent* due to its wide coverage on the media.

2.3. The German Political System

Germany is a federal country and comprises 16 states (*Länder)*.[5] Each state *(Land)* has its own constitution, elects its own parliament and creates its own government. Matters of national importance, such as foreign affairs, defense, or citizenship, are competence of the federal parliament *(Bundestag)* and government, while each state parliament *(Landtag)* has full autonomy on various subject matters, such as education, culture, police, or the press.[6] Elections in federal states occur at different times and with different electoral laws. A typical legislative period lasts five years.[7] Parliamentary debates in each state parliaments occur regularly, and, on average, 1.9 parliamentary sessions take place each month in each state.

The main political forces in the German political systems in the period analyzed in this paper, 2000-2008, consist of a left-leaning social democratic bloc, represented by the Social Democrats (SPD) and the Green Party (GRÜNE), and a right-leaning conservative bloc, represented by the Christian-Democratic Union (CDU) with its sister Bavarian denomination (CSU), and the Liberal Party (FDP).

3. Measuring Polarization in Parliamentary Debates: Data, Methods, and Descriptive Statistics

In this section, I describe the data sources, the methods used to compute the polarization measure, and report descriptive statistics of the main data sources.

[5] An exhaustive description of the German political system is outside the scope of this paper. In this section, I highlight only the features that are most relevant for the scope of this paper.
[6] A further category, which includes subjects such as environment, nature protection or land use, are jointly regulated by the federal and state parliaments. Interested readers may find the complete list of competences in https://www.bpb.de/medien/189018/Foederalismus.pdf.
[7] Except for Bremen, where legislatures last four years.

3.1. Parliamentary Debates of the German States

The main source of data for this paper consists of parliamentary debates of the 16 German states for the period January 2000 – August 2008. Data limitations, discussed in more detail in Appendix B, prevent me from using data before the year 2000. The financial crisis that began in September 2008 serves as the cutoff point for my analysis, as it may have influenced the salience of numerous topics. Parliamentary debates constitute the preferred data source to measure the polarization for a variety of reasons. First, they convey timely and abundant information about MPs' opinions as opposed to voting patterns of member of parliaments, an alternative approach that has often been used to measure polarization in the US.[8] Second, parliamentary debates are a crucial way through which politicians obtain visibility in the media (Maltzman and Sigelman 1996; Tresch 2009; Salmond 2014) and express their views (Proksch and Slapin 2012), thus making them relevant for the policymaking process. As some scholars have argued, MPs use parliamentary speeches mainly as an act of position-taking rather than to persuade opponents or win political arguments (Proksch and Slapin 2015). Parliamentary debates are therefore particularly suited to study the extent to which MPs' policy positions evolve over time and across parties.

The federal structure of Germany also provides the ideal setting for this study. First, with respect to other studies using national parliamentary debates (Peterson and Spirling 2018; Goet 2019; Salla 2020; Fiva, Nedregård, and Øien Henning 2022), this setting yields a much higher density of parliamentary debates, which is crucial to overcome the high-dimensionality issue inherent to text data (Gentzkow, Shapiro, and Taddy 2019). Second, state elections do not occur at the same time, which ensures that my results are not driven by the idiosyncratic distance from upcoming elections or political leanings.

I obtained the entire population of parliamentary debates for the period of interest of each German state as PDF documents by scraping each state's official website.[9] I then created a dataset that includes all speeches from the 16 German states for the period

[8] Ideological positions measured with roll call-based approaches tend not to be informative in parliamentary systems such as Germany (Spirling and McLean 2007; Peterson and Spirling 2018). Further drawbacks of roll-call analyses include the selection of votes subject to roll call and their ability to capture only high levels of inter-party disagreement (Proksch and Slapin 2015). As noted in Slagter and Loewenberg (2007), roll-call votes occurred frequently in the German Bundestag in the period between 1949 and 1957, a period characterized by considerable party differences, whereas their frequency plunged between 1957 and 1983, which reflected an inter-party consensus on many issues and a desire to avoid public scrutiny.

[9] Parliamentary debates of Saarland are not available in the official website for the period considered in this analysis. Nevertheless, these debates were made available for my research upon request.

2000-2008. This process involved several steps to extract the data from the gathered documents in order to record all the relevant information contained in the documents, such as the speeches, name and role of the speaker, party affiliation, interruptions, state, and date in which the debates occurred.[10] I complemented this dataset with information about the date of the latest and next election and with the shares obtained by the two major German parties, the CDU/CSU and SPD, in the latest election in each state.[11]

The unit of analysis is a speech as recorded in the parliamentary debates. I consider a speech the continuous utterance issued by the same person. During a speech, speakers are often interrupted by remarks of other speakers, applauses etc. Such interruptions are excluded from the speeches.

3.2. Topic Classification of Parliamentary Debates

I classify the topic of each speech in the parliamentary debates. This step enables me to compute topic-specific measures of polarization, which are crucial for both my identification strategy and to overcome issues inherent to measuring polarization in parliamentary debates that I describe in the next subsection.

I use a combination of supervised and unsupervised machine learning algorithms.[12] First, I used a supervised machine learning to classify speeches in a binary way: whether they are about education or not. This approach requires a subset of manually-labelled speeches which are used to train the model. For this purpose, I obtained a set of 3,346 manually-labelled speeches with which I trained a supervised machine learning model. I then used the best-performing model, a Logistic Classifier, to make the out-of-sample predictions for the entire corpus. I report the in-sample performance of the classifier in Table C1 and the result of a validation exercise in Table C2. Both tables suggest a reliable classification. The share of speeches classified as being about education is 8.9%, or 18,703 speeches.

I then used an unsupervised machine learning model, namely the correlated topic model (CTM), to classify the topic of all the speeches that were classified as *not* being about education in the previous step. The key hyperparameter to tune the CTM is the number of topics. A CTM with 30 topics provided good results in terms of

---

[10] Interested readers can find detailed information about the process of gathering the necessary documents, extracting text from the documents and creating a unified corpus of parliamentary debates in Appendix B.

[11] I retrieved these data from Metawahl, an open-source project that collects data of all German elections (last accessed 7th November 2022).

[12] Interest readers may find a detailed description of this classification task in Appendix C. In this section, I describe only the most relevant aspects.

interpretability of the topics. I then aggregated the estimated topics into 11 topics of similar size as the education topic classified in the previous step. I report the estimated topics, most representative words and the assigned label in Table C3.

3.3. Measuring Polarization in Parliamentary Debates

Measuring polarization in parliamentary debates is challenging. A fundamental problem is that the words used in legislative speeches are a function of both the topic of the debate and the position of the speaker (Lauderdale and Herzog 2016).[13] Hence, the use of different words across MPs from different parties might be mistakenly attributed to polarization when in fact it might be due to MPs discussing different topics. Previous work has dealt with this issue by, for example, limiting the analysis to a single legislative act (Herzog and Benoit 2015), by comparing speeches only within a specific debate (Lauderdale and Herzog 2016) , or, conversely, combining speeches over many debates for each legislator or party, assuming that the resulting documents contain the same mixture of topics (e.g., Giannetti and Laver 2005; Proksch and Slapin 2010). I tackle this issue in a novel way. I first classify the topic of each speech in the parliamentary debates, as explained in the previous subsection. I then compute polarization within each topic, which allows me to isolate the different words used by MPs due to polarization from the different words used due to MPs talking about different topics.

A further issue concerns the finite-sample bias that arises because the pool of words a speaker can choose from is large relative to the total amount of speech we observe (Gentzkow, Shapiro, and Taddy 2019). This implies that many words are used only by MPs of one party just by chance, and naïve estimators might interpret such differences as evidence of polarization. I tackle this issue excluding words that are mentioned in less than 10 speeches within a topic from the computation of the polarization measure.[14] This ensures that rare words, which are more likely to be uttered only by MPs of one party just by chance, do not drive my measure of polarization.[15] Germany's 16 state

[13] In fact, there are even more sources of variation in word usage. In descending order of importance, these are: language, style, topic, and position (Lauderdale and Herzog 2016). Given the context, it is safe to assume that language and style are reasonably homogeneous within parliamentary debates.

[14] In Section 5.5, I show that results are robust to different thresholds.

[15] I elaborate more on this intuition and formalize it in footnote 29 in Section 6. To account for the finite-sample bias, Gentzkow, Shapiro, and Taddy (2019) specify a multinomial model of speech that they estimate through a penalized Lasso model to compute an accurate measure of polarization. Their approach, however, does not account for the different topics MPs address in their speeches.

parliaments provide a substantial amount of parliamentary debates in each topic, enabling this approach.

To compute polarization, I first perform standard preprocessing steps such as removal of stopwords, punctuation and numbers. I then transform each speech d about topic s into an adjusted term-frequency vector according to the following topic-specific term-frequency inverse-document frequency ($tf-idf$) formula:

$$tf-idf_{ds} \equiv \frac{c_{dw}}{\sum_{k\in d} c_{dk}} \times \ln\left(\frac{D^S}{\sum_{n\in D^S} \mathbb{I}(c_{nw}>0)}\right), \quad (1)$$

where the relative term frequency of each term w in speech d ($c_{dw}/\sum_{k\in d} c_{dk}$) is weighted by the natural logarithm of the inverse frequency of the term w in all the speeches D in topic $s \in S$ ($\ln(D^S/\sum_{n\in D^S} \mathbb{I}(c_{nw} > 0))$).

Compared to the standard $tf-idf$ transformation of a document, the topic-specific $tf-idf$ that I use also upweights words that occur frequently in a document, but downweights words that appear often in many documents *about the same topic.* Hence, words that are mentioned often only in a specific topic will receive less weight, thus alleviating the risk of attributing the use of different words to polarization when in fact it is due to speakers discussing different topics. Further, I also drop rare words, which mitigates the finite-sample bias mentioned previously.[16]

I define polarization as the extent to which opinions on a topic are opposed. Assuming that politicians use different words to express different opinions, the more polarized the speeches, the less similar they are. I therefore use a straightforward measure of text (dis)similarity: the opposite of the cosine similarity between the vector representation of each speech d in topic s and state-legislative period cell l and the vector representation of all the speeches from a benchmark party r, the CDU/CSU,[17] in topic s and state-legislative period cell l. Formally the polarization of a speech is computed as follows:

$$polarization_{dsl} \equiv -\frac{\sum_i A_{idsl}\bar{B}_{isl}}{\sqrt{\sum_i A_{idsl}^2}\sqrt{\sum_i \bar{B}_{isl}^2}}, \quad (2)$$

[16] Formally, for a threshold τ, only words w for which $\sum_{n\in D^S} \mathbb{I}(c_{nw} > 0) > \tau$ are kept. In Section 5.5, I show that results are robust to different thresholds used at this stage.

[17] I show in Section 5.5 that the results are robust to using different benchmark parties or factions.

where A_{idsl} is the $tf-idf$ vector representation of speech d in topic s and state-legislative period cell l, and $\bar{B}_{isl}$ is the average of the vector representation of all the speeches by MPs that belong to benchmark party r in topic s and state-legislative period cell l:

$$\bar{B}_{isl} \equiv \frac{\sum_{p\in r} B_{ipsl}}{\sum_{p\in r} \mathbb{I}(B_{psl})} \tag{3}$$

Thus, $\bar{B}_{isl}$ captures the "average" speech of a benchmark party r in a specific topic, state, and legislative period. The less similar a speech is to $\bar{B}_{isl}$, the larger the polarization measure. In the next subsection, I provide evidence to validate the polarization measure by showing that it captures differences across parties in word use.

3.4. Descriptive Statistics

The entire dataset consists of 622,946 speeches. I drop all the speeches by the President of each state parliament, 327,498 speeches, as these are strictly procedural and not informative of the political debates. I also drop all speeches with less than 100 words, namely 100,816 speeches, as these are too short to be reliably classified among different topics. The resulting sample consists of 210,006 speeches, and descriptive statistics of the dataset are reported in Table 1. The average length of a speech is 663.6 words. The share of speeches by ministers of each state parliament is 24%. The share of speeches issued by members of the main center-right party, CDU/CSU, is 34%, while the share for main center-left party, SPD, is 27%. These parties represent the main political forces in Germany and are the only parties that have been part of each German state parliament in the entire period considered. The second tier of political forces in the German landscape in this period is represented by the Green party and the FDP, the liberal party, with a share of speeches of 14% and 11%, respectively. Speeches from these four parties make up 86% of the entire corpus of parliamentary debates. The remaining 14% of speeches are uttered by member of minor parties, none of which reaches the threshold of 10% of all the speeches in the corpus.[18]

---

[18] Among these minor parties, the most relevant is the Left party, with its various denomination over time and states (DIE LINKE, the current one or, previously, Linksfraktion, Linkspartei.PDS, PDS, REGENBOGEN), whose share of speeches is 8.8%. Other minor parties include a series of extreme right parties (DVU, DVU-FL, FDVP, NPD, PRO, REP, Ronald-Schill-Fraktion), whose combined share of speeches in the corpus is 2.9%. The remaining 2.1% of speeches are uttered by MPs of local parties (0.96%), MPs whose party could not be identified (0.8%), or without a political affiliation (0.37%).

TABLE 1—DESCRIPTIVE STATISTICS

	Mean (1)	SD (2)	Min/Max (3)
Word Count	663.57	(622.52)	100.0-17503.0
Share CDU/CSU	0.34	(0.47)	0.0-1.0
Share SPD	0.27	(0.44)	0.0-1.0
Share GREENS	0.14	(0.34)	0.0-1.0
Share FDP	0.11	(0.32)	0.0-1.0
Share Ministers	0.24	(0.42)	0.0-1.0
Share Gov. Speeches	0.53	(0.50)	0.0-1.0
Share Education Speeches	0.09	(0.28)	0.0-1.0
# Observations		210,006	
# States		16	
# Parl. Sessions		3,277	

Notes: Descriptive statistics of speeches from parliamentary debates. The share of speeches is reported separately only for parties for which the total number of speeches is larger than 10% of the entire corpus of speeches. The number of observations coincides with the number of speeches.

The PISA shock had a substantial impact on the public debate about education. In Figure 1, I report the share of respondents from a representative survey of the German population that indicate education as the most or second most important problem in Germany. Such share increased dramatically after the PISA shock. In the two years prior to the PISA shock, only 2.6% of respondents indicated education as the most or second most important problem in Germany on average. This share more than doubled after the PISA shock: on average, 5.7% of respondents indicated education as the most or second most important problem in Germany in the seven years after the PISA shock. The release of the results of the subsequent PISA study, three years later, had a similar impact on the public opinion. It is also interesting to note that the PISA shock triggered an upward trend in the importance of education, as it never reverted to its pre-shock level in the seven years after the shock.

A similar pattern emerges when looking at parliamentary debates. I report the share of speeches about education and the number of times that “PISA” was mentioned in parliamentary debates in Figure 2. This figure clearly depicts the “tsunami-like” impact of the release of the first PISA results on the political debate about education. The share of speeches about education increased by 1.8 percentage points after the PISA shock. This effect translates into a 22% increase with respect to the pre-shock share of 7.3% and is statistically significant (see Table A3). In the first six months after the PISA shock, the term “PISA” was mentioned more than 2,000 times in parliamentary debates. Overall, “PISA” was mentioned almost 11,000 times after the PISA shock. These figures substantiate the claim that the salience of education increased dramatically because of the PISA shock. I will analyze the impact of this exogenously induced increase in salience of education on the polarization of political debates in Section 5.

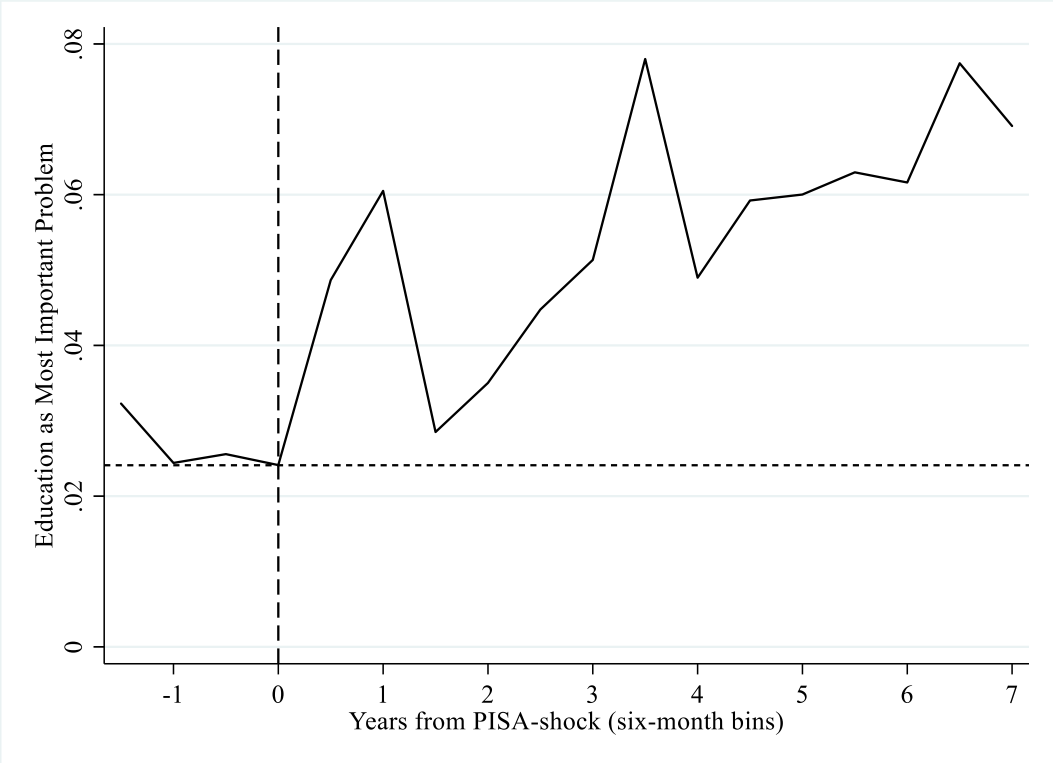


FIGURE 1. EDUCATION AS MOST IMPORTANT PROBLEM

Notes: Data source: Politbarometer (Forschungsgruppe Wahlen 2019). The y-axis reports the share of respondents that indicated education as the most or second most important problem in Germany. The x-axis reports the distance (in years) from the PISA shock, which occurred on the 4th of December 2001. Data are aggregated into six-month bins.

I report the estimated topics and size in Figure A1. With roughly 9% of the speeches, education is a mid-sized topic in the corpus, whereas the largest topic concern economic issues and the lawmaking process. For about 5% of the speeches no clear topic could be identified, and I therefore assigned the label "Other" to this topic. In Figure A2, I report the speeches' topic size by state. No major difference in the distribution of topics across states can be observed. The education topic, in green, appears to be quite homogenous across states.

Finally, I report evidence to validate the polarization measure in Figure A3. As expected, the polarization measure aggregated at the party level is much lower for the CDU/CSU when the CDU/CSU is used as the benchmark party in the left panel. By this measure, the average speech from a member of the SPD is 0.48 SD more polarized than the average speech from a member of the CDU/CSU party. Similarly, the average polarization measure for members of the CDU/CSU is much larger when the SPD is used as the benchmark party. This suggests that the polarization measure captures meaningful differences in word use across MPs of different parties.

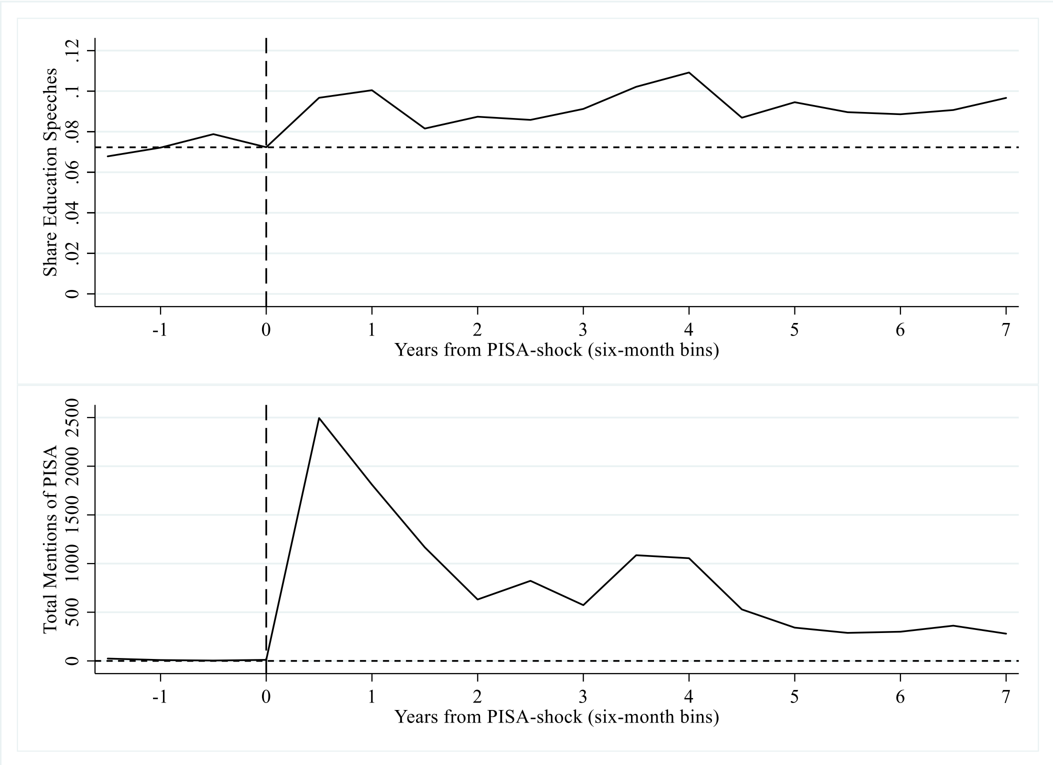


FIGURE 2. THE “TSUNAMI”-LIKE IMPACT OF PISA

Notes: The figure reports the share of education speeches in parliamentary debates in the upper panel and the total number of mentions of the term “PISA” in parliamentary debates in the lower panel. The x-axis reports the distance (in years) from the PISA shock, which occurred on the 4th of December 2001. Data are aggregated into six-month bins.

3.5. Additional Data Sources: State-Specific PISA Results and Bills

The performance in the PISA 2000 reading test of each German state is reported in Table A1. State-specific results were released on the 25th of June 2002, almost seven months after the PISA shock. There is a large heterogeneity in the performance. The average score of the best performing German state, Bayern, is 62% of a standard deviation higher than the lowest performing state, Bremen. Such difference corresponds to the distance between the best performing state in the reading test of PISA 2000, Finland, and Germany, whose performance was well below the OECD average. It is also important to note that the state-specific results of Berlin and Hamburg were not released due to low participation rates.

I also use data from the "Pattern of Lawmaking in the German Länder" dataset (Stecker, Kachel, and Paasch 2021), which comprises all 16,610 bills that have been initiated in the 16 German state parliaments between 1990 and 2020. The dataset contains a wealth of information regarding the bills. For the purpose of my analysis, the main variables of interest are the initial date on which the bill was proposed, the status of each bill—whether the bill was adopted, rejected or other—, the topic of each bill, which has been manually coded, and a German state identifier. For consistency with the rest of the analysis, I use data for the period January 2000 - August 2008 and report their descriptive statistics in Table A2. Specifically, I report the total number of proposed bills by each topic as defined in the dataset, the share of bills by each topic, as well as the total number of bills by their status. With 525 proposed bills, education is the largest topic in the dataset and covers 10% of the bills. Reassuringly, this share is very close to the share of speeches about education in parliamentary debates (roughly 9%, see Figure A1). Since topics in the law-making dataset were manually coded, this improves the credibility of the classification task I carried out for this project. The other topics in the law-making dataset are more narrowly defined than those that I estimated for the parliamentary debates, which makes the comparison less meaningful. During the period of interest, 5,356 bills were initiated. Out of all the initiated bills, 4,116 (76.9%) have been adopted, while 821 (15.3%) were rejected. Thus, the large majority of proposed bills have been adopted, which reflects the fact that bills tend to be initiated by governing parties who have the political power to adopt them.[19] The status of the remaining 419 (7.8%) bills, labelled as "Other", includes exceptional cases of bills which have been withdrawn, discontinued, adjourned etc.

4. Empirical Strategy

Estimating the causal effect of the salience of a topic on polarization in parliamentary debates requires exogenous variation in the salience of a topic. As argued in the previous sections, the PISA shock in Germany led to an exogenous increase in the salience of the education topic, which rules out issues of reverse causation. It therefore provides an ideal setting to study its impact on the polarization of parliamentary debates.

[19] 92% of bills that are eventually adopted have been initiated by governing parties, whereas 99% of rejected bills have been initiated by opposition parties. Hence, there is almost a complete overlap between adopted (rejected) bills and bills initiated by governing (opposition) parties. A minority of bills have been initiated by bipartisan coalitions (3.7%), and they have been adopted in 97% of the cases.

I exploit the fact that the PISA shock affected a single topic, education, to implement a difference-in-differences strategy.[20] The key idea is that speeches about unaffected topics act as counterfactuals for speeches about education that occurred after the PISA shock, thus accounting for underlying trends in polarization of parliamentary debates and for time-invariant differences among polarization in different topics. I therefore estimate the following equation:

$$y_{islt,r\neq b} = \theta_s + \alpha PostPISA_t + \beta PostPISA \times Ed_{st} + \gamma' X_{ilt,r\neq b} + \sigma_l + \varepsilon_{islt,r\neq b} \tag{4}$$

The outcome variable $y_{islt,r\neq b}$ denotes the polarization between speech i by member of party r and all the speeches of benchmark party $r = b$ in topic s and state-legislative period cell l at time t. Speeches from the benchmark parties are therefore omitted from the analysis. θ_s denotes topic fixed effects, which account for differences in level of polarization across topics and the dummy variable $PostPISA_t$ accounts for differences before and after the PISA shock, which occurred on the 4th of December 2001. The interaction term, $PostPISA \times Ed_{st}$ takes value one if a speech occurred after the PISA shock and if it is about education. In this setup, the parameter of interest β can be estimated by means of the two-way fixed effects estimator (TWFE), which accounts for time-invariant differences between treated and untreated units.

$X_{ilt,r\neq b}$ is a vector of speech, state, and time specific controls, such as the length of the speech i, the shares of the two main parties, CDU/CSU and SPD, at time t in state-legislative period cell l, whether the speech i is given by a member of a governing party, is given by a minister, distance from the next election in state-legislative period cell l at time t, year and party fixed effects. As shown in Section 5.1, the length of speech i plays an important role as a control, since it is negatively correlated with the polarization measure and including it in the regression causes a substantial increase in the R^2 of the model. However, including it as a control is potentially problematic if the PISA shock also affected the verbosity of the speeches. At the same time, it ensures that the results are not driven by an increase or decrease in the verbosity of the speeches. σ_l denotes state-legislative period fixed effects; that account for differences in the level of polarization across state-legislative period cells. $\varepsilon_{islt,r\neq b}$ is the idiosyncratic error. I standardize the polarization measure to have mean zero and standard deviation one to

[20] I show in Section 5.5 that the PISA shock did not affect polarization in other topics.

interpret the estimated coefficients in terms of standard deviation. I cluster standard errors at the state level throughout the paper.

The identification strategy rests on the assumption of parallel trends of the treated and untreated units. In this application, this means that the polarization in education debates would have trended similarly to other topics in the absence of the PISA shock. While this assumption is not directly testable, I exploit the availability of multiple time periods before the shock to show the absence of different pre-trends between education and other topics in Section 5.1.

Another identifying assumption is that the effect of the PISA shock affected the polarization of education debates through topic salience. The effect could also be driven by the negative results of German students revealed by the PISA study rather than the salience of the education topic. I tackle this issue in Section 5.2, where I exploit the fact that in June 2002, six months after the PISA shock, the results for all but two German states were published. Despite the large heterogeneities in the performance of German states and the fact that results were not published for two states, I show that the effect of the PISA shock on polarization was homogenous across German states.[21] Further, the low performance of German students in international standardized assessment was already shown by the TIMSS study in 1995, but this event was largely ignored by the German media (see Section 2.1). Hence, the results revealed by the PISA study were not completely new to German MPs. This further corroborates the assumption that the effect on polarization was driven by the salience induced by the PISA shock rather than the information revealed by the PISA study.

5. Results

5.1. Main Results

I report evidence of the validity of the parallel-trends assumption using an event-study design in Figure 3, where I interact the dummy variable indicating whether a speech is about education and year fixed effects. The figure does not show diverging trends in the period prior to the PISA shock, and I cannot reject the null hypothesis of pre-event effects being zero, thus suggesting that polarization in political debates about education and other topics were following the same trend before the shock. Conversely, the test of

[21] Note that in this setting heterogenous treatment effects would not bias the TWFE estimator. As the recent literature on difference-in-differences methods noted, heterogenous treatment effects can bias the TWFE estimator if units are treated at different point in times (see Roth et al. 2022 for a review of this literature), which is not the case in this setting.

post-event effects being jointly null is largely rejected. It can also be noted that the impact of the PISA shock on polarization seemingly fades out over time and that polarization reverts to its pre-shock level only about six years after the shock.

I provide further evidence of the validity of the parallel trend assumption in Figure A4 and Figure A5. In Figure A4, I report point estimates of the pre-trends by interacting the education dummy with six-month bins instead of yearly bins to increase the number of pre-trend point estimates. Even in this specification, I do not find significantly different pre-trends between education and other topics, although standard errors become substantially larger. In Figure A5, I show the dynamic of polarization in all the estimated topics in the period of interest net of the controls and fixed effects described in Equation (4). The polarization of education debates clearly increased after the shock, while no similar patterns can be detected for other topics. In sum, both figures provide evidence in favor of the validity of the parallel trend assumption.

I report the estimates of Equation (4) in Table 2. The magnitude of the impact varies between 8% of a SD in the most parsimonious specification in column 1, and 11.1% SD in a specification that also includes state-legislative period, party, and year fixed effects in column 2. All coefficients are statistically significant. The main difference between column 2 and column 3 concerns the inclusion of the length of a speech as a control, which causes a decrease in the estimated coefficient to 8.8% SD. At the same time, including it more than doubles the R^2 of the model. I therefore prefer the most restrictive specification in column 3, which should be therefore considered as a conservative estimate.[22] An increase of 8.8% SD in polarization is equivalent to 18% of the polarization between the main center-right (CDU/CSU) and center-left (SPD) parties.[23] Overall, these results show that the PISA shock had a substantial and persistent impact on the political debates about education.

[22] If the verbosity of the speeches was affected by the PISA shock, the inclusion of length of speeches as a control could be problematic. In fact, I find weak evidence that the PISA-shock caused speeches in education to become roughly 5.5% shorter by substituting the logarithm of length of speech as the outcome variable in Equation (4). Nonetheless, including length of speech as a control ensures that the impact of the PISA shock on polarization occurred above and beyond the verbosity of the speeches.

[23] The share is the absolute value of the estimated coefficient (0.088) divided by the difference between the polarization measure for the CDU/CSU and the SPD (0.48) when the CDU/CSU is used as the benchmark party reported in Figure A3.

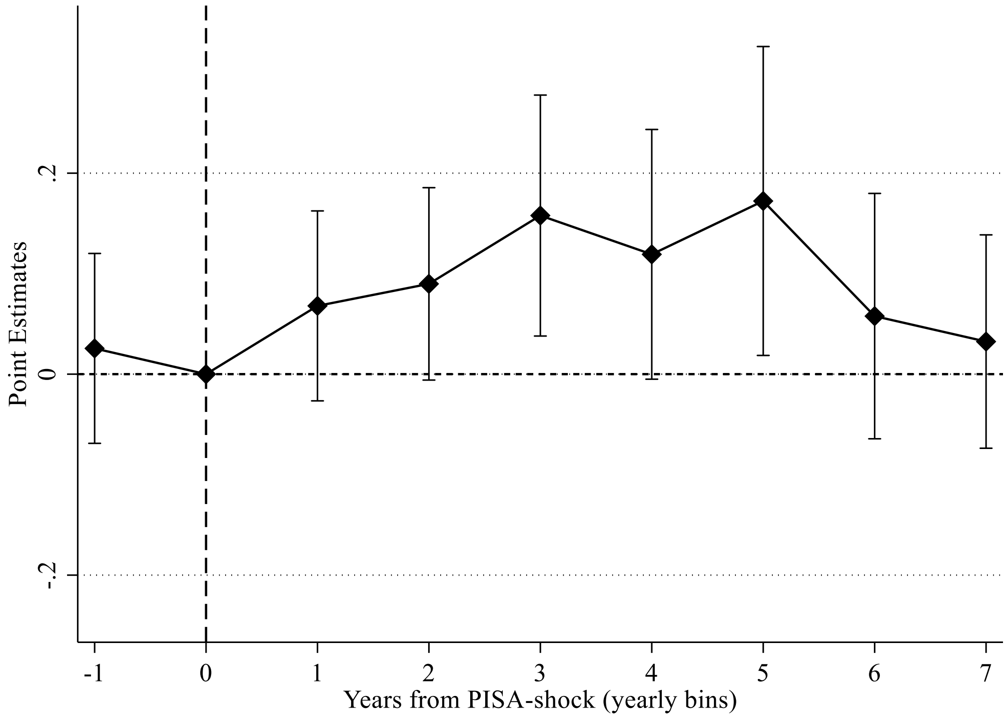


FIGURE 3. THE IMPACT OF THE PISA SCHOCK ON POLARIZATION IN EDUCATION DEBATES: EVENT-STUDY GRAPH

Notes: Event-study estimates of the impact of the PISA shock on polarization with 95% confidence intervals. The estimated equation takes the following form: $y_{islt,r\neq b} = \theta_s + \alpha PostPISA_t + \sum_{\tau\in(-1,7),\tau\neq 0} \beta_\tau Ed_s \times \mathbb{I}(t+\tau) + \gamma' X_{ilt,r\neq b} + \sigma_l + \varepsilon_{islt,r\neq b}$. This event-study setup takes the same form as Equation (4), but instead of pooling years before and after shock, I interact the education dummy with an indicator variable for each year $(t+\tau)$. I label t the year before the PISA shock, which I consider the reference year. The pre-shock period covers the years $t-1$ and t, while the post-shock period covers the years from $t+1$ to $t+7$ (until August 2008). Standard errors have been clustered at the state level. The dependent variable is the standardized polarization. The x-axis reports the distance (in years) from the PISA shock, which occurred on the 4th of December 2001. The year prior to the PISA shock is the excluded category. The p-values of the joint hypothesis tests of zero pre- and post-event effects are 0.603 and 0.001, respectively.

5.2. State-Specific Heterogeneities

As argued in Section 4, a possible concern regarding identification strategy is that the impact of the PISA shock on polarization is not due to the increased salience of education. The new information revealed by the PISA study about the low performance of German students might have also caused the increase in polarization. To test this hypothesis, I leverage the fact that the initial PISA shock, which occurred on 4th of December 2001, was followed by a state-specific PISA shock on the 26th of June 2002. On this date, German state-specific results for all but two states were released and revealed large heterogeneities in the performance of German states (reported in Table A1).

TABLE 2—PISA SHOCK AND POLITICAL POLARIZATION IN EDUCATION DEBATES - DIFFERENCE-IN-DIFFERENCES

	(1)	(2)	(3)
PISA shock × Education	0.080*	0.111**	0.088**
	(0.042)	*(0.042)*	*(0.038)*
Topic FE	Yes	Yes	Yes
State-Legislative Period FE	No	Yes	Yes
Party, Year FE	No	Yes	Yes
Controls	No	No	Yes
R^2	0.148	0.260	0.535
Observations	137,820	137,820	137,820

Note: Difference-in-differences estimate of the impact of the PISA shock on the polarization of education speeches. The dependent variable is the standardized polarization with CDU/CSU as the benchmark party. All regressions include topic fixed effects and a dummy for whether the speeches occurred after the PISA shock. Controls include the length of a speech, the shares obtained at the latest state election by the two main parties, CDU/CSU and SPD, and a dummy variable for whether a speech is given by a minister or a state secretary, and if the MPs belongs to the governing coalition, and the distance from the next election. The data include all parliamentary debates from January 2000 till August 2008. Standard errors (in parentheses) have been clustered at the state level. *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

I therefore investigate whether the impact of the PISA shock differed with respect to the actual performance of each state. To this purpose, I first create an additional treatment variable ("PISA shock (State)") to capture whether a speech occurred after the state-specific PISA shock of the 26th of June 2002. I then interact this variable with a "PISA-Published-Score" dummy variable, that takes value one for the states of Berlin and Hamburg for which the PISA state results were not published. Second, I interact the dummy "PISA shock (State)" with a set of dummies that capture whether each state's performance was in the lower, middle, or upper tercile of the distribution of performance of German states. I further explore this hypothesis by interacting the "PISA shock (State)" treatment with the performance of each German state.[24] Results in Table 3 show that the impact of the PISA shock was homogenous not only with respect to whether state specific results were published or not (column 2), but also with respect to the actual performance of each German state (column 3 and 4).[25]

[24] PISA tests three subjects: math, reading, and science. In each wave, PISA has a special focus on one of the three subjects. Since reading was the focus of PISA in the first wave, I use the performance in reading (reported in column 1, Table A1); using math or science performance leads to the same results (table not shown).

[25] Another potentially interesting dimension of heterogeneity concerns former West and East German states. Again, I do not find statistically significant differences in the impact of the PISA shock on the polarization of education debates in former West and East German states (results not shown).

The lack of sizable heterogeneity across states also emerges in Table A3, which shows little differences of the impact of the PISA shock on the share of education speeches. The only marginally significant difference emerges with respect to the states for which the state-specific results were not published, namely Hamburg and Berlin (column 3). The share of speeches about education increased slightly less after the PISA shock in these states. Overall, these results suggest that the salience of the topic, rather than the actual performance of the students revealed by the PISA study affected the polarization of the debates.

5.3. Heterogeneity by Party

I explore which parties contributed the most to the increase in polarization in Table 4. It is worth reminding that, since the benchmark party is the CDU/CSU, party interactions capture the polarization of each party with respect to the CDU/CSU. Results show that the increase in polarization is driven by a cleavage between the two main parties, the CDU/CSU and the SPD. In fact, the interaction between treatment dummy and the SPD dummy is positive and reaches a 10% level of statistical significance in column 5, where all the interactions are included. Conversely, the FDP and the Green Party do not appear to contribute substantially to the increase in polarization. To corroborate these results, I repeat the analysis using the polarization measure between left- and right-wing parties, which allows me to include speeches from all the parties in the regression.[26] I report results from this specification in Table A4. Again, the increase in polarization seems to be driven by the CDU/CSU and SPD, whose associated coefficient is positive in column 1 and 2, and reaches statistical significance when all the interactions are included in column 5. Results from this section are compatible with a cleavage theory framework, where the main center-right and center-left parties exploit the increased salience of education induced by the PISA shock to amplify their ideologically distinctiveness.

[26] I show in Section 5.5 (Table 6, column 2) that the main results are essentially the same when using this measure of polarization. The advantage of this measure is that for each speech of members of right-(left-) wing parties, all the speeches from the members of the left-(right-)wing parties in the same topic, state, and legislative period are used as benchmark speeches.

TABLE 3—HETEROGENEITIES BY STATE-SPECIFIC PERFORMANCE

	(1)	(2)	(3)	(4)
PISA shock *(Federal)* × Education	0.058	0.058	0.048	0.049
	(0.038)	*(0.038)*	*(0.033)*	*(0.033)*
PISA shock *(State)* × Education	0.033	0.033	0.033	0.033
	(0.036)	*(0.036)*	*(0.024)*	*(0.024)*
PISA shock *(State)* × Education × PISA-Published-Score		0.000		
		(0.017)		
PISA shock *(State)* × Education × Med. Perf.			0.009	
			(0.017)	
PISA shock *(State)* × Education × High Perf.			-0.022	
			(0.026)	
PISA shock *(State)* × Education × PISA Perf./ 100				-0.070
				(0.057)
Topics, State-Legisl. Period, Party, Year FE	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
R^2	0.628	0.620	0.635	0.635
Observations	137,820	137,820	119,462	119,462

Note: Difference-in-differences estimate of the impact of the PISA shock on the polarization of education speeches. The dependent variable is the standardized polarization with CDU/CSU as the benchmark party. All regressions include a dummy for whether the speeches occurred after the federal or state PISA shock, topic, state-legislative period, party, and year fixed effects. Controls include the length of a speech, the shares obtained at the latest state election by the two main parties, CDU/CSU and SPD, and a dummy variable for whether a speech is given by a minister or a state secretary, and if the MPs belongs to the governing coalition, and the distance from the next election. The variable, "PISA shock *(Federal)*" is a dummy variable which takes value one if a speech occurred after 4th December 2001. The variable "PISA shock *(State)*" is a dummy variable that takes value one if a speech occurred after 26th June 2002. The medium performance variable takes value one if the performance of the respective state is in the middle tercile, while high performance takes value one if the performance is in the upper tercile. In column 3, the omitted category is the lower tercile. The variable "PISA Performance" is the performance of each state in the PISA 2000 reading test. The data include all parliamentary debates from January 2000 till August 2008. Standard errors (in parentheses) have been clustered at the state level. *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

TABLE 4—HETEROGENEITIES BY PARTY

	(1)	(2)	(3)	(4)	(5)
PISA shock × Education	0.088**	0.056	0.086*	0.094*	-0.003
	(0.038)	*(0.038)*	*(0.044)*	*(0.048)*	*(0.087)*
PISA shock × Education × SPD		0.071			0.130*
		(0.051)			*(0.062)*
PISA shock × Education × FDP			0.015		0.103
			(0.059)		*(0.101)*
PISA shock × Education × GREENS				-0.028	0.069
				(0.076)	*(0.103)*
Topics, State-Legisl. Period, Party, Year FE	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes
R^2	0.535	0.535	0.535	0.535	0.535
Observations	137,820	137,820	137,820	137,820	137,820

Notes: Difference-in-differences estimate of the impact of the PISA shock on the polarization of education speeches. The dependent variable is the standardized polarization with CDU/CSU as the benchmark party. All regressions include a dummy for whether the speeches occurred after the PISA shock, topic, state-legislative period, party, and year fixed effects. Controls include the length of a speech, the shares obtained at the latest state election by the two main parties, CDU/CSU and SPD, and a dummy variable for whether a speech is given by a minister or a state secretary, and if the MPs belongs to the governing coalition, and the distance from the next election. The data include all parliamentary debates from January 2000 till August 2008. Standard errors (in parentheses) have been clustered at the state level. *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

5.4. The Impact of the PISA Shock on the Number of Bills

Topic salience might also affect the number of bills discussed in parliaments. Bills are the main output of parliaments and, therefore, they represent a proxy of parliaments' productivity. I investigate whether MPs respond to the salience of a topic by increasing their effort concerning such topic. I use data on law-making in German state parliaments collected by Stecker, Kachel, and Paasch (2021), which allows me to implement essentially the same identification strategy described in Section 4, where my treated group consists of bills about education proposed after the PISA shock. The outcome variable is the logarithm of the number of bills in each topic and state in a six-month bin. I assign bills to the six-month bin in which the bill was proposed. Estimated coefficients from this log-linear model can be therefore interpreted as percentage changes in the number of bills. I report results for the overall number of proposed bills, as well as separately for rejected and adopted bills.

TABLE 5—PISA SHOCK AND BILLS ABOUT EDUCATION -DIFFERENCE-IN-DIFFERENCES

	All Bills			Adopted	Rejected
	(1)	(2)	(3)	(4)	(5)
PISA shock × Education	0.201**	0.211**	0.162**	0.121	0.261*
	(0.079)	*(0.077)*	*(0.073)*	*(0.081)*	*(0.130)*
Topic FE	Yes	Yes	Yes	Yes	Yes
State FE	No	Yes	Yes	Yes	Yes
Year FE	No	No	Yes	Yes	Yes
Observations	2,931	2,931	2,931	2,510	547
R^2	0.240	0.250	0.255	0.238	0.307

Notes: Difference-in-differences estimate of the impact of the PISA shock on the number of bills about education in all German states. The dependent variable is the natural logarithm of the total number of bills about each topic in a state, topic, six-month-bin cell. Each observation corresponds to a state-topic-six-month-bin cell. All regressions include a dummy for whether the speeches occurred after the PISA shock. In columns 1-3, all the bills are used, regardless of their status. In column 4 and 5, I restrict the sample to accepted bills and rejected bills, respectively. The data include all proposed bills from January 2000 till August 2008. Standard errors (in parentheses) have been clustered at the state level. *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

Results indicate a 16.2-21.1% increase in the total number of proposed bills about education because of the PISA shock (column 1-3). This suggests that MPs indeed put more effort into this topic. At a closer look, the effect is driven by the number of rejected bills (column 5). As discussed in Section 3.5, virtually all rejected bills are proposed by the opposition. It is therefore possible that that MPs in the opposition strategically propose more bills in a salient topic to signal to voters their effort in this topic, despite the very low chances of such bills being adopted.

It is interesting to note that in this context both the polarization and the number of bills in education increased. This is surprising given that polarization has often been linked with gridlocks in parliament (Jones 2001; Binder 2004; Lapinski 2008; McCarty, Poole, and Rosenthal 2016), which hinders the law-making process. My results suggest that an increase in polarization and vibrant law-making can coexist, although they do not offer a clear interpretation of the relationship between these two concepts, which lies outside the scope of this paper.

5.5. Robustness Checks

A first concern about the validity of my main results regards the polarization measure. The choice of a benchmark party for the computation of the polarization measure entails a certain degree of arbitrariness. I therefore compute alternative measures of polarizations by varying the benchmark parties of faction.[27] In Table 6, I show that using speeches of different parties or factions as a benchmark does not appreciably alter the main results. I only report the results using the most restrictive specification, which controls for topic, state-legislative period, party and year fixed effects, as well as the controls described in Equation (4).

In column 1, I report the results obtained using speeches of the SPD as the benchmark party. In column 2 and 3, I do not use a single party as the benchmark to compute the polarization measure. Instead, I compute the cosine similarity between each speech of right (left)-wing parties and all the speeches from the left (right)-wing parties within the same topic, state, and legislative period. I report results for this polarization measure in column 2. Similar to column 2, in column 3 I report the results obtained computing the cosine similarity between each speech from a governing party and all the speeches from parties in the opposition, and vice versa.

Differently from the specification using a single party as the benchmark corpus to compute the polarization measure, these specifications allow me to include all speeches in the regressions, since an appropriate benchmark exists for all speeches. This comes at the cost of using as a benchmark a corpus of speeches which is more heterogenous, as it comprises speeches of different parties. In fact, despite the substantial increases in the

[27] Practically, the measures differ because of the different speeches used in Equation (**3** to compute the "average" speech $\bar{B}_{isl}$ against which the polarization measure is computed.

number of observations, the standard errors in column 2 and 3 do not decrease appreciably, possibly due to the heterogeneity of the benchmark corpus.

Regardless of the benchmark party or faction chosen, the results are remarkably robust. The coefficient estimated in the main specification and reported in Table 2, column 3 (0.088), lies between the coefficient obtained when using the SPD as the benchmark party in column 1 (0.90), and the coefficient estimated when using the left/right-wing polarization measure in column 2 (0.086).

TABLE 6—MAIN RESULTS WITH DIFFERENT BENCHMARK PARTIES OR FACTIONS

	SPD (1)	Left/Right (2)	Gov./Opp. (3)
PISA shock × Education	0.090^{**}	0.086^{**}	0.077^{*}
	(0.038)	*(0.033)*	*(0.037)*
Topic, State-Legisl. Period, Party, Year FE	Yes	Yes	Yes
Controls	Yes	Yes	Yes
R^2	0.535	0.553	0.560
Observations	152,464	205,160	209,459

Notes: Difference-in-differences estimate of the impact of the PISA shock on the polarization of education speeches. The dependent variable is the standardized polarization with SPD as the benchmark party in column 1, speeches of parties of the opposite wing (i.e., left or right) as the benchmark corpus in column 2, and speeches of opposite the coalition (i.e., governing or opposition) as the benchmark corpus in column 3. All regressions include a dummy for whether the speeches occurred after the PISA shock, topic fixed effects. Controls include the length of a speech, the shares obtained at the latest state election by the two main parties, CDU/CSU and SPD, and a dummy variable for whether a speech is given by a minister or a state secretary, and if the MPs belongs to the governing coalition, and the distance from the next election. The data include all parliamentary debates from January 2000 till August 2008. Standard errors (in parentheses) have been clustered at the state level. *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

Second, I conduct a robustness check to ensure that the observed effect is due to the PISA shock and not to other events, such as the release of results of subsequent PISA studies, which occur every three years, or other events that might affect the polarization in the counterfactual topics. To this purpose, I restrict the sample to speeches that occurred two years before and two years after the shock. This specification also ensures a balanced sample size of the pre- and post-shock period. I report estimates of this specification in Table A5. Results are very similar to the main results in Table 2 and, if anything, larger in magnitude in the preferred estimated in column 3 (0.096 SD).

Another concern regards the number of topics. As discussed in section 3 and, more in detail, in Appendix C, the number of topics chosen depends on a variety of factors, such as the size of the corpus, previous knowledge of the researcher, and the downstream task one wants to achieve. To balance the interpretability of the topics and ensure that topics were of similar size to the education topics, I estimated a CTM with 30 topics, which I then aggregated into 11 topics. In Table A6, I report the results obtained estimating a CTM with the following number of topics: 9, 10, 11, 13, and 15. To maximize the transparency of this exercise, I also do not aggregate the topics as done previously. Results suggest that using a number of topics similar to the number of aggregated topics that I use does not affect the main results substantially. This suggests that neither the number of topics chosen, nor the aggregation step are driving the results in the preferred specification.

I further corroborate my findings by conducting a placebo test, where I test the effect of the PISA shock on the polarization of the other topics. Had the PISA shock also affected the polarization of other topics, estimates might be biased, since the affected topics would not constitute an appropriate counterfactual. I report results in Figure A6. In each row I report the coefficient obtained interacting the PISA shock dummy with a dummy for the topic indicated in each row along with 95% confidence intervals. I report only the estimated coefficient obtained using the preferred specification described in Equation (4, which includes topic, state-legislative period, party and year fixed effects and controls for the length of each speech and distance from elections.

In the first row, I report the results from the main specification, where education is the treated topic. In the subsequent rows, I report the coefficients from the placebo exercise. Besides the coefficient for education, only the coefficients for the topic "Local Politics" and "Social Welfare, Healthcare and Equality" reach the 10% threshold of statistical significance, while the other coefficients do not reach any conventional threshold of statistical significance. I cannot entirely rule out that these effects are due to the PISA shock, but other events occurred in the period 2000-2008 might also have affected the polarization in such topics. As results reported in Table A7 show, when restricting the placebo exercise to a symmetric time window around the shock (2000-2004), the placebo coefficients in column 2 and 3 for "Local Politics" and "Social Welfare, Healthcare and Equality", respectively, are not statistically significant anymore, whereas the coefficient for education in column 1 remains positive and statistically significant. This suggests that the change in polarization in these topics is due to other events that occurred after the PISA shock.

To a large extent, results from the placebo exercise alleviate the concern that the effect of the PISA shock on the polarization of education debates is biased by the simultaneous impact of the PISA shock on the polarization of other topics. I further address such concern with a leave-one-topic-out exercise, where I iteratively estimate Equation (4) by dropping one of the counterfactual topics at each iteration. This robustness check shows that results are not driven by any topic in the counterfactual group that might have been affected by the PISA shock or other events. I report results in Figure A7 with 95% confidence intervals. In the first row, I include all the topics and coefficient is the therefore same as the coefficient reported in Table 2, column 3. In the subsequent rows, I report the estimated coefficient obtained by dropping the topic indicated in each row. The estimated coefficients are relatively stable and remain statistically significant regardless of which topic is excluded from the estimation sample.

Finally, I report results obtained by changing the threshold above which words are kept to compute the polarization measure in Table A8. As mentioned in Section 3.3, to avoid the sample-finite bias in the polarization measure I only use words that are mentioned in at least ten speeches. This ensures that rare words, which are more likely to be uttered only by MPs of one party just by chance, do not drive the polarization measure. I have therefore computed alternative measures of the main polarization measure with the CDU/CSU as the benchmark party obtained by imposing more restrictive thresholds. In column 1-3, I report results obtained by using words that are mentioned in at least 20, 30 or 40 speeches within a topic. In columns 4-6, I report results obtained using words that are mentioned in at least 2%, 2.5%, and 5% of speeches within a topic. Results are robust to these different thresholds.

6. Polarizing Issues in Education Debates

6.1. Polarization Score

In the previous section, I showed that polarization in education debates increased as a consequence of the PISA shock. I have shown that the effect was mainly driven by the two main center-right and center-left parties, the CDU/CSU and SPD, respectively. In this section, I provide suggestive evidence on what are the most polarizing issues in education debates. I focus on the two main parties that drove the increase in polarization, the CDU/CSU and the SPD, and on debates about education. For each term in $w \in W$, where W denotes the vocabulary of terms uttered by MPs of either the CDU/CSU or SPD in debates about education, I develop a polarization score $p(w)$, which is defined as follows:

$$p(w) \equiv \frac{f\ (w_{CDU}) - f\ (w_{SPD})}{f\ (w_{CDU}) + f\ (w_{SPD})} \times \ln\big(f\ (w_{CDU}) + f\ (w_{SPD})\big) \quad (5)$$

where $f\ (w_{CDU})$ ($f\ (w_{SPD})$) denotes the total number of times the term w is mentioned by the CDU/CSU (SPD). The first part of the score varies between -1 and 1, where 1 (-1) indicates terms that have only been mentioned by MPs that belong to the CDU/CSU (SPD). This part is weighted by the natural logarithm of the total number of times the term w has been mentioned by either the CDU/CSU or the SPD.

The rationale for this polarization score is simple. In absolute value, terms that display high polarization scores are those that *(i)* tend to be mentioned more often by one party and *(ii)* are mentioned often. Terms that are uttered the same number of times by both parties will get a polarization score of 0. Terms that are uttered more often by one party but are relatively infrequent will be pushed toward zero.[28] Hence, the polarization score of rare terms, for which there is a higher probability that they are uttered only or mostly by one party just by chance,[29] will be pushed toward 0.

The polarization score closely mirrors the polarization measure that I use throughout the analysis, which is based on the cosine similarity between a corpus of speeches from a benchmark party and speeches from the other parties. Similar to the cosine similarity, the polarization score depends on both the frequency with which one term is used by one party and on its absolute frequency. This ensures that terms that have high polarizing scores are also those that drive the polarization measure in the education debates.

6.2. Polarizing Issues in Education

I focus on the 10,000 most frequent terms uttered by either member of the CDU/CSU and SPD in education speeches, after removing uninformative terms such as stopwords, names, and numbers. This ensures that these terms are unlikely to obtain large

---

[28] Note that, in the extreme case where a term w is mentioned only once and, therefore, is mentioned only by one party, $p(w) = 0$, since $f\ (w_{CDU}) + f\ (w_{SPD}) = 1$ and $\ln(1) = 0$.

[29] To formalize this intuition, let us consider a generic term w, for which $f\ (w_{CDU}) + f\ (w_{SPD}) = N$, with $N \in \mathbb{N}, N > 0$. Let us assume that w is a neutral term, i.e., that each realization of $f\ (w_{CDU})$, without loss of generality, is equally likely: $f\ (w_{CDU})\ \sim U(0, N)$ and $E[f\ (w_{CDU})] = \frac{N}{2}$. Hence, the probability that the term w is uttered by only one party is: $P(f\ (w_{CDU}) = 0) + P(f\ (w_{CDU}) = N) = \frac{1}{N} + \frac{1}{N} = 2/N$. Thus, the smaller the N, or, equivalently, the rarer the term w, the higher the probability that w is uttered only by MPs of one party just by chance, since $P(f\ (w_{CDU}) = n)$ strictly decreases in N. More generally, for an arbitrarily small $n \in \mathbb{N}$, $P(f\ (w_{CDU}) \leq n) = n/N$. Thus, the larger the N, the lower the probability that terms are mentioned primarily by MPs of one party just by chance.

polarization scores just by chance (see footnote 26). On average, these terms are mentioned 254,7 times and 50% of the terms are mentioned at least 88 times. The minimum frequency of a term is 35. I report the distribution of the polarization score in Figure A8. I rescaled the polarization score to have a zero mean and divided it by $\max(|p(w)|)$, so that $-1 \leq p(w) \leq 1$. The distribution is quite concentrated around the mean; the standard deviation of the distribution is 0.16 and for 50% of the terms $|p(w)| \leq 0.1$.

I focus on the 250 terms with the largest polarization score for each party, or the top 5% of polarizing terms. For the CDU/CSU (SPD), these terms lie in the black (red)-shaded area in Figure A8. These terms have a polarization score $|p(w)| \geq 0.32$. I display the 250 terms with the highest CDU/CSU (SPD) polarization score in Figure A9 (Figure A10), translated in English (Panel (a)) and in the original language (German, Panel (b)).

A variety of findings emerge from the most polarizing terms. I primarily focus on those issues that were particularly relevant in the aftermath of the PISA shock. As mentioned in Section 2.1, the three most important issues that emerged from the PISA shock were: developing a monitoring system with common educational standards and central examination, expanding "all-day school" offers, and reforming the tracking system. It is interesting to notice that terms related to these issues can be found among the most polarizing terms in Figure A9 and Figure A10.

Concerning a monitoring system with common educational standards and central examination, the term "state exams" (*"Landesprüfungen"*) appears as a strongly polarized term favored by the CDU/CSU. Conversely, the term "learning assessments" ("*Lernstandserhebungen"*) is a strongly polarized term favored by the SPD. This terminology suggests polarized views on the ways to monitor the education systems: while the CDU/CSU favored a testing regime of central state exams, which are typically high-stake exams for students, the SPD seemingly favored a testing regime aimed at monitoring student achievement in a low-stake environment. As a matter of fact, state exams, in particular those at the end of high school in Germany, have been introduced in most states in the years after the PISA shock. In 2000, only 7 states had a central upper secondary school leaving examination ("*Zentralabitur"*).[30] From 2004 to 2008, this examination was gradually rolled out to all German states except Rhineland-

[30] The states are: Bavaria, Baden-Württemberg, Mecklenburg-Vorpommern, Saarland, Saxony, Saxony-Anhalt.

Palatinate (Helbig and Nikolai 2015). At the same time, a plan to establish a new set of common standards was also implemented. In 2004, the Institute for Quality Development in Education (IQB) was created to develop math, reading, writing and foreign-language standards and accompanying tests (Neumann, Fischer, and Kauertz 2010; OECD 2011).

A second issue concerns the expansion of all-day schooling. Again, a term linked to this concept can be found among strongly polarized terms: the term "all-day elementary school" ("*Ganztagsgrundschule"*) is a polarized term favored by the SPD. This picture is in line with the account by Kuhlmann and Tillmann (2009), according to which the SPD was promoting the expansion of the all-day schooling offer since the end of 2001, as it considered it an effective policy to improve equal opportunities for students.[31] Conversely, the CDU/CSU considered all-day schooling as a threat to the family and, therefore, hindered its expansion for a long time. Despite the different stances toward this issue, the offer of all day-schools in Germany was rapidly expanded thanks to large subsidies granted by the German national government through the investment program "*Future, Education, and Care* (2003-2009)" (IZBB).[32]

A third relevant issue was the tracking system. Until 2000, the large majority of students in all German states were tracked into three main different ability schools at the age of 10. Given the large educational inequality highlighted by the PISA shock across German students with different socio-economic and, in particular, migration backgrounds, strong arguments were made against the existing three-tiered early-tracking system. Again, terms related to this concept can be easily found among the most polarized terms. For example, the terms "sorting" or "selecting" ("*sortieren"* and "*aussortieren*", respectively) are terms typically used by the SPD. Conversely, the term "comprehensive school" ("*Einheitsschule*") is a strongly polarized term used by the CDU/CSU. A comprehensive school is opposed to the three-tier school system typical of Germany. While some states enacted reforms to reduce the segregation induced by

[31] Relatedly, the term "equality of opportunity ("*Chancengleichheit"*) also features among the strongly polarized terms favored by the SPD.
[32] Detail of the program can be found at https://www.ganztagsschulen.org/de/service/izbb-programm/das-investitionsprogramm-zukunft-bildung-und-betreuung-izbb (last accessed: 16 December 2022).

the early-tracking system,[33] the distinction between three hierarchical school tracks has been mostly left intact (Henninges, Traini, and Kleinert 2019).

Overall, this section offers suggestive evidence on three possible issues that might have led to an increase in the polarization of education debates. It is interesting to note that polarization in two of these topics, the monitoring system and the all-day schooling, was accompanied by important and substantial reforms on these issues. Conversely, the tracking system was not largely addressed by the reforms.

7. Conclusion

The rise of polarization observed in many democracies has fueled a lively debate on the causes of such phenomenon. While research on the determinants of polarization in the electorate abounds, much less is known about what drives polarization in political speech. In this paper, I shed light on topic salience as a possible determinant of polarization in parliamentary debates. I find that the sharp increase in the salience of education induced by the PISA shock in Germany had a strong and long-lasting impact on the polarization of debates about education. I do not find heterogeneities across states, despite considerable differences in the performance of students in different states revealed by the PISA shock. These results lend support to the cleavage theory of political behavior as opposed to a convergence toward the median platform, whereby MPs amplify their ideological distinctiveness with respect to a salient topic. The results are robust to different measures of polarization, to different numbers of topics in the counterfactual group and to a variety of robustness checks. I also find an increase in the number of proposed bills about education, which is driven by rejected bills. The simultaneous increase in polarization and in the number of proposed bills is an interesting pattern which challenges previous findings in the literature, that have often linked high polarization with gridlocks in parliament.

I also provide suggestive evidence that issues related to developing a monitoring system with common educational standards and central examination, expanding all-day school offers, and reforming the tracking system led to the increase in polarization of education debates. While the first two topics were subject to substantial reforms in the

[33] For example, some states have merged the two lower-level tracks ("*Realschule*" and "*Hauptschule*") into one school, called regional schools ("*Regionalschulen*") (Davoli and Entorf 2018). Despite this trend toward a two-tier education system, the issue of access to the academic track ("*Gymnasium"*), which constitutes the main route to a tertiary degree, has not been addressed (Henninges, Traini, and Kleinert 2019).

aftermath of the PISA shock, the tracking system was not largely addressed. This provides further evidence that polarization in parliamentary debates and the legislative process do not necessarily overlap.

References

Artelt, Cordula, Jürgen Baumert, Eckhard Klieme, Michael Neubrand, Manfred Prenzel, Ulrich Schiefele, Wolfgang Schneider, Klaus-Jürgen Tillmann, and Manfred (Hrsg.) Weiß. 2002. *PISA 2000 – Die Länder Der Bundesrepublik Deutschland Im Vergleich:* Leske + Budrich, Opladen. Accessed February 18, 2022. https://www.grundschulpaedagogik.uni-bremen.de/archiv/pisa/PISA2000Lesen/PISA-E.pdf.

Ash, Elliott, Massimo Morelli, and Richard van Weelden. 2017. "Elections and Divisiveness: Theory and Evidence." *The Journal of Politics* 79 (4): 1268–85. https://doi.org/10.1086/692587.

Autor, David, David Dorn, Gordon Hanson, and Kaveh Majlesi. 2020. "Importing Political Polarization? The Electoral Consequences of Rising Trade Exposure." *American Economic Review* 110 (10): 3139–83. https://doi.org/10.1257/aer.20170011.

Bergbauer, Annika B., Eric A. Hanushek, and Ludger Woessmann. 2021. "Testing." *J. Human Resources*, 0520-10886R1. https://doi.org/10.3368/jhr.0520-10886R1.

Berger, Lara M. 2022. "How Digital Media Markets Amplify News Sentiment.".

Bieber, Tonia, and Kerstin Martens. 2011. "The OECD PISA Study as a Soft Power in Education? Lessons from Switzerland and the US." *European Journal of Education* 46 (1): 101–16.

Binder, Sarah A. 2004. *Stalemate: Causes and Consequences of Legislative Gridlock:* Brookings Institution Press.

Blei, David M., and John D. Lafferty. 2007. "A Correlated Topic Model of Science." *Ann. Appl. Stat.* 1 (1). https://doi.org/10.1214/07-aoas114.

Bleich, Erik, and A. Maurits van der Veen. 2021. "Media Portrayals of Muslims: A Comparative Sentiment Analysis of American Newspapers, 1996–2015." *Politics, Groups, and Identities* 9 (1): 20–39. https://doi.org/10.1080/21565503.2018.1531770.

Bonica, Adam. 2013. "Ideology and Interests in the Political Marketplace." *Am J Political Science* 57 (2): 294–311. https://doi.org/10.1111/ajps.12014.

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer. 2022. "Salience." *Annu. Rev. Econ.* 14 (1): 521–44. https://doi.org/10.1146/annurev-economics-051520-011616.

Bos, Wilfried, and T. Neville Postlethwaite. 2002. "Internationale Schulleistungsforschung: Ihre Entwicklungen Und Folgen Für Die Deutsche Bildungslandschaft." In *Leistungsmessungen in Schulen.*, edited by Franz E. (H.) Weinert. 2nd: Weinheim: Beltz.

Boxell, Levi, Matthew Gentzkow, and Jesse M. Shapiro. 2022. "Cross-Country Trends in Affective Polarization." *The Review of Economics and Statistics*, 1–60. https://doi.org/10.1162/rest_a_01160.

Breakspear, Simon. 2012. "The Policy Impact of PISA: An Exploration of the Normative Effects of International Benchmarking in School System Performance." *OECD Education Working Papers.* https://doi.org/10.1787/5k9fdfqffr28-en.

Canen, Nathan, Chad Kendall, and Francesco Trebbi. 2020a. "Political Parties as Drivers of U.S. Polarization: 1927-2018." https://doi.org/10.3386/w28296.

Canen, Nathan, Chad Kendall, and Francesco Trebbi. 2020b. "Unbundling Polarization." *ECTA* 88 (3): 1197–1233. https://doi.org/10.3982/ECTA16756.

Davoli, Maddalena, and Horst Entorf. 2018. "The PISA Shock, Socioeconomic Inequality, and School Reforms in Germany." *IZA Policy Paper*.

DellaVigna, S., and E. Kaplan. 2007. "The Fox News Effect: Media Bias and Voting." *The Quarterly Journal of Economics* 122 (3): 1187–1234. https://doi.org/10.1162/qjec.122.3.1187.

Der Spiegel. 2001. "Sind Deutsche Schüler Doof?" *Der Spiegel*, December 13, 2001. https://www.spiegel.de/lebenundlernen/schule/die-pisa-analyse-sind-deutsche-schueler-doof-a-172357.html.

Downs, Anthony. 1957. *An Economic Theory of Democracy:* New York: Harper and Brothers.

Ebenrett, Heinz J., Dieter Hansen, and Klaus J. . Puzicha. 2003. "Verlust Von Humankapital in Regionen Mit Hoher Arbeitslosigkeit." *Aus Politik und Zeitgeschichte* B6 (7).

Ertl, Hubert. 2006. "Educational Standards and the Changing Discourse on Education: The Reception and Consequences of the PISA Study in Germany." *Oxford Review of Education* 32 (5): 619–34. http://www.jstor.org/stable/4618685.

FAZ. 2001. "Katastrophales Zeugnis Für Deutsche Schüler." *FAZ*, December 2, 2001.

Figlio, David, and Susanna Loeb. 2011. "School Accountability." In *Handbook of the Economics of Education, Vol. 3,*. Vol. 3, edited by Eric A. Hanushek, Stephen Machin, and Ludger Woessmann, 383–421. Handbook of the Economics of Education. Amsterdam: North Holland: Elsevier.

Fiva, Jon H., Oda Nedregård, and Øien Henning. 2022. "Polarization in Party-Centered Environments: Evidence from Parliamentary Debates." *Working Paper*. https://www.jon.fiva.no/docs/FivaNedregardOien.pdf. Accessed 3rd, February 2022.

Forschungsgruppe Wahlen, Mannheim. 2019. "Politbarometer - Gesamtkumulation.".

Funke, Manuel, Moritz Schularick, and Christoph Trebesch. 2016. "Going to Extremes: Politics After Financial Crises, 1870–2014." *European Economic Review* 88 (3): 227–60. https://doi.org/10.1016/j.euroecorev.2016.03.006.

Galey-Horn, Sarah, Sarah Reckhow, Joseph J. Ferrare, and Lorien Jasny. 2020. “Building Consensus: Idea Brokerage in Teacher Policy Networks.” *American Educational Research Journal* 57 (2): 872–905. https://doi.org/10.3102/0002831219872738.

Gentzkow, Matthew, Jesse M. Shapiro, and Matt Taddy. 2019. “Measuring Group Differences in High-Dimensional Choices: Method and Application to Congressional Speech.” *ECTA* 87 (4): 1307–40. https://doi.org/10.3982/ECTA16566.

Giannetti, Daniela, and Michael Laver. 2005. “Policy Positions and Jobs in the Government.” *European Journal of Political Research* 44 (1): 91–120. https://doi.org/10.1111/j.1475-6765.2005.00220.x.

Global Education Monitoring Report Team. 2017. *Accountability in Education: Meeting Our Commitments; Global Education Monitoring Report, 2017/8:* Paris: UNESCO.

Goet, Niels D. 2019. “Measuring Polarization with Text Analysis: Evidence from the UK House of Commons, 1811–2015.” *Polit. Anal.* 27 (4): 518–39. https://doi.org/10.1017/pan.2019.2.

Grek, Sotiria. 2009. “Governing by Numbers: The PISA ‘Effect’ in Europe.” *Journal of Education Policy* 24 (1): 23–37. https://doi.org/10.1080/02680930802412669.

Gruber, Karl H. 2006. “The German "PISA-Shock": Some Aspects of the Extraordinary Impact of the OECD's PISA Studyon the German Education System.” In *Cross-National Attraction in Education: Accounts from England and Germany*, edited by Hubert Ertl, 195–208: Symposium Books.

Helbig, Marcel, and Rita Nikolai. 2015. *Die Unvergleichbaren: Der Wandel Der Schulsysteme in Den Deutschen Bundesländern Seit 1949.* Bad Heilbrunn: Verlag Julius Klinkhardt.

Henninges, Miriam, Claudia Traini, and Corinna Kleinert. 2019. “Tracking and Sorting in the German Educational System.” *Leibniz Institute for Educational Trajectories (LIfBi) Working Paper No* 83.

Herzog, Alexander, and Kenneth Benoit. 2015. “The Most Unkindest Cuts: Speaker Selection and Expressed Government Dissent During Economic Crisis.” *The Journal of Politics* 77 (4): 1157–75. https://doi.org/10.1086/682670.

Hopmann, Stefan, Gertrude Brinek, and Martin Retzl. 2007. *PISA Zufolge PISA: Hält PISA, Was Es Verspricht? = PISA According to PISA : Does PISA Keep What It Promises?* Schulpädagogik und pädagogische Psychologie Bd. 6. Wien: Lit.

Iyengar, Shanto, Yphtach Lelkes, Matthew Levendusky, Neil Malhotra, and Sean J. Westwood. 2019. “The Origins and Consequences of Affective Polarization in the United States.” *Annual Review of Political Science* 22 (1): 129–46. https://doi.org/10.1146/annurev-polisci-051117-073034.

Jensen, Jacob, Ethan Kaplan, Suresh Naidu, and Wilse-Samson Laurence. 2012. “Political Polarization and the Dynamics of Political Language: Evidence from 130 Years of Partisan Speech.” *Brookings Papers on Economic Activity*, 1–81. Accessed 9th, February 2022.

Jones, David R. 2001. "Party Polarization and Legislative Gridlock." *Political Research Quarterly* 54 (1): 125–41. https://doi.org/10.1177/106591290105400107.

Kayser, Mark A., and Michael Peress. 2021. "Does the Media Cover the Economy Accurately? An Analysis of Sixteen Developed Democracies." *QJPS* 16 (1): 1–33. https://doi.org/10.1561/100.00019098.

Kuhlmann, Christian, and Klaus-Jürgen Tillmann. 2009. "Mehr Ganztagsschulen Als Konsequenz Aus PISA? Bildungspolitische Diskurse Und Entwicklungen in Den Jahren 2000 Bis 2003." In *Ganztagsschule Als Symbolische Konstruktion*, edited by Fritz-Ulrich Kolbe, Sabine Reh, Till-Sebastian Idel, Bettina Fritzsche, and Kerstin Rabenstein, 23–45. Wiesbaden: VS Verlag für Sozialwissenschaften.

Lapinski, John S. 2008. "Policy Substance and Performance in American Lawmaking, 1877–1994." *Am J Political Science* 52 (2): 235–51. https://doi.org/10.1111/j.1540-5907.2008.00310.x.

Lauderdale, Benjamin E., and Alexander Herzog. 2016. "Measuring Political Positions from Legislative Speech." *Polit. Anal.* 24 (3): 374–94. https://doi.org/10.1093/pan/mpw017.

Levendusky, Matthew S. 2013. "Why Do Partisan Media Polarize Viewers?" *American Journal of Political Science* 57 (3): 611–23. https://doi.org/10.1111/ajps.12008.

Lewandowsky, Marcel, Julia Schwanholz, Christoph Leonhardt, and Andreas Blätte. 2022. "New Parties, Populism, and Parliamentary Polarization: Evidence from Plenary Debates in the German Bundestag." In *The Palgrave Handbook of Populism*. Vol. 36, edited by Michael Oswald, 611–27. Cham: Springer International Publishing.

Lipset, Seymour M., and Stein Rokkan. 1967. *Party Systems and Voter Alignments: Cross-National Perspectives:* New York: The Free Press.

Lyon, Melissa Arnold, and Matthew A. Kraft. 2021. "Elevating Education in Politics: How Teacher Strikes Shape Congressional Election Campaigns.".

Maltzman, Forrest, and Lee Sigelman. 1996. "The Politics of Talk: Unconstrained Floor Time in the U.S. House of Representatives." *The Journal of Politics* 58 (3): 819–30. https://doi.org/10.2307/2960448.

Martens, Kerstin, and Dennis Niemann. 2013. "When Do Numbers Count? The Differential Impact of the PISA Rating and Ranking on Education Policy in Germany and the US." *German Politics* 22 (3): 314–32. https://doi.org/10.1080/09644008.2013.794455.

McCarty, Nolan, Keith T. Poole, and Howard Rosenthal. 2016. *Polarized America, Second Edition: The Dance of Ideology and Unequal Riches:* MIT Press.

McDonnell, Lorraine M., and M. Stephen Weatherford. 2013. "Organized Interests and the Common Core." *Educational Researcher* 42 (9): 488–97. https://doi.org/10.3102/0013189X13512676.

Mian, Atif, Amir Sufi, and Francesco Trebbi. 2014. "Resolving Debt Overhang: Political Constraints in the Aftermath of Financial Crises." *American Economic Journal: Macroeconomics* 6 (2): 1–28. https://doi.org/10.1257/mac.6.2.1.

Neumann, Knut, Hans E. Fischer, and Alexander Kauertz. 2010. "FROM PISA to EDUCATIONAL STANDARDS: THE IMPACT of LARGE-SCALE ASSESSMENTS on SCIENCE EDUCATION in GERMANY." *Int J of Sci and Math Educ* 8 (3): 545–63. https://doi.org/10.1007/s10763-010-9206-7.

OECD. 2011. *Lessons from PISA for the United States.* Strong Performers and Successful Reformers in Education. Paris: OECD Publishing.

Peterson, Andrew, and Arthur Spirling. 2018. "Classification Accuracy as a Substantive Quantity of Interest: Measuring Polarization in Westminster Systems." *Polit. Anal.* 26 (1): 120–28. https://doi.org/10.1017/pan.2017.39.

Prior, Markus. 2013. "Media and Political Polarization." *Annu. Rev. Polit. Sci.* 16 (1): 101–27. https://doi.org/10.1146/annurev-polisci-100711-135242.

Proksch, Sven-Oliver, and Jonathan B. Slapin. 2012. "Institutional Foundations of Legislative Speech." *American Journal of Political Science* 56 (3): 520–37. https://doi.org/10.1111/j.1540-5907.2011.00565.x.

Proksch, Sven-Oliver, and Jonathan B. Slapin, eds. 2015. *The Politics of Parliamentary Debate:* Cambridge University Press.

Rinne, Risto, Johanna Kallo, and Sanna Hokka. 2004. "Too Eager to Comply? OECD Education Policies and the Finnish Response." *European Educational Research Journal* 3 (2): 454–85. https://doi.org/10.2304/eerj.2004.3.2.3.

Roberts, Margaret E., Brandon M. Stewart, and Dustin Tingley. 2019. "Stm : An R Package for Structural Topic Models." *J. Stat. Soft.* 91 (2). https://doi.org/10.18637/jss.v091.i02.

Ryu, Jung S. 1982. "Public Affairs and Sensationalism in Local TV News Programs." *Journalism Quarterly* 59 (1): 74–137. https://doi.org/10.1177/107769908205900111.

Salla, Simola. 2020. "A Century of Partisanship in Finnish Political Speech." *Working Paper*. https://sites.google.com/site/sallasimolaecon/home/. Accessed 3rd, February 2022.

Salmond, Rob. 2014. "Parliamentary Question Times: How Legislative Accountability Mechanisms Affect Mass Political Engagement." *The Journal of Legislative Studies* 20 (3): 321–41. https://doi.org/10.1080/13572334.2014.895121.

Settle, Jaime E. 2018. *Frenemies: How Social Media Polarizes America:* Cambridge University Press.

Slagter, Tracy Hoffman, and G. Loewenberg. 2007. "The Persistence of Procedural Consensus in the German Bundestag." *Unpublished Manuscript*.

Sloane, Peter F.E., and Bernadette Dilger. 2005. "The Competence Clash – Dilemmata Bei Der Übertragung Des 'Konzepts Der Nationalen Bildungsstandards' Auf Die Berufliche Bildung." *Berufs- und Wirtschaftspaedagogik*. https://www.bwpat.de/ausgabe8/sloane_dilger_bwpat8.shtml. Accessed February 18, 2022.

Soroka, Stuart, Mark Daku, Dan Hiaeshutter-Rice, Lauren Guggenheim, and Josh Pasek. 2018. "Negativity and Positivity Biases in Economic News Coverage: Traditional Versus Social Media." *Communication Research* 45 (7): 1078–98. https://doi.org/10.1177/0093650217725870.

Spirling, Arthur, and Iain McLean. 2007. "UK OC OK? Interpreting Optimal Classification Scores for the U.K. House of Commons." *Polit. Anal.* 15 (1): 85–96. https://doi.org/10.1093/pan/mpl009.

Stecker, Christian, Jannis Kachel, and Jana Paasch. 2021. "Patterns of Lawmaking in the German Länder.".

Sunstein, Cass R. 2018. *#Republic: Divided Democracy in the Age of Social Media:* Princeton University Press.

TAZ. 2001. "Fast in Jeder Hinsicht Ein Desaster." *TAZ*, December 5, 2001. https://taz.de/!1137421/.

Tillmann, Klaus-Jürgen. 2004. "Was ist eigentlich neu an PISA? Zum Verhältnis von erziehungswissenschaftlicher Forschung, öffentlicher Diskussion und bildungspolitischem Handeln." https://doi.org/10.25656/01:2571.

Tresch, Anke. 2009. "Politicians in the Media: Determinants of Legislators' Presence and Prominence in Swiss Newspapers." *The International Journal of Press/Politics* 14 (1): 67–90. https://doi.org/10.1177/1940161208323266.

Waldow, Florian. 2009. "What PISA Did and Did Not Do: Germany After the 'PISA-Shock'." *European Educational Research Journal* 8 (3): 476–83. https://doi.org/10.2304/eerj.2009.8.3.476.

West, Martin R., and Ludger Woessmann, eds. 2021. *Public Opinion and the Political Economy of Education Policy Around the World.* CESifo seminar series. Cambridge Massachusetts: The MIT Press.

Woessmann, Ludger. 2009. *School Accountability, Autonomy and Choice Around the World.* Cheltenham: Edward Elgar.

Appendix A – Tables and Figures

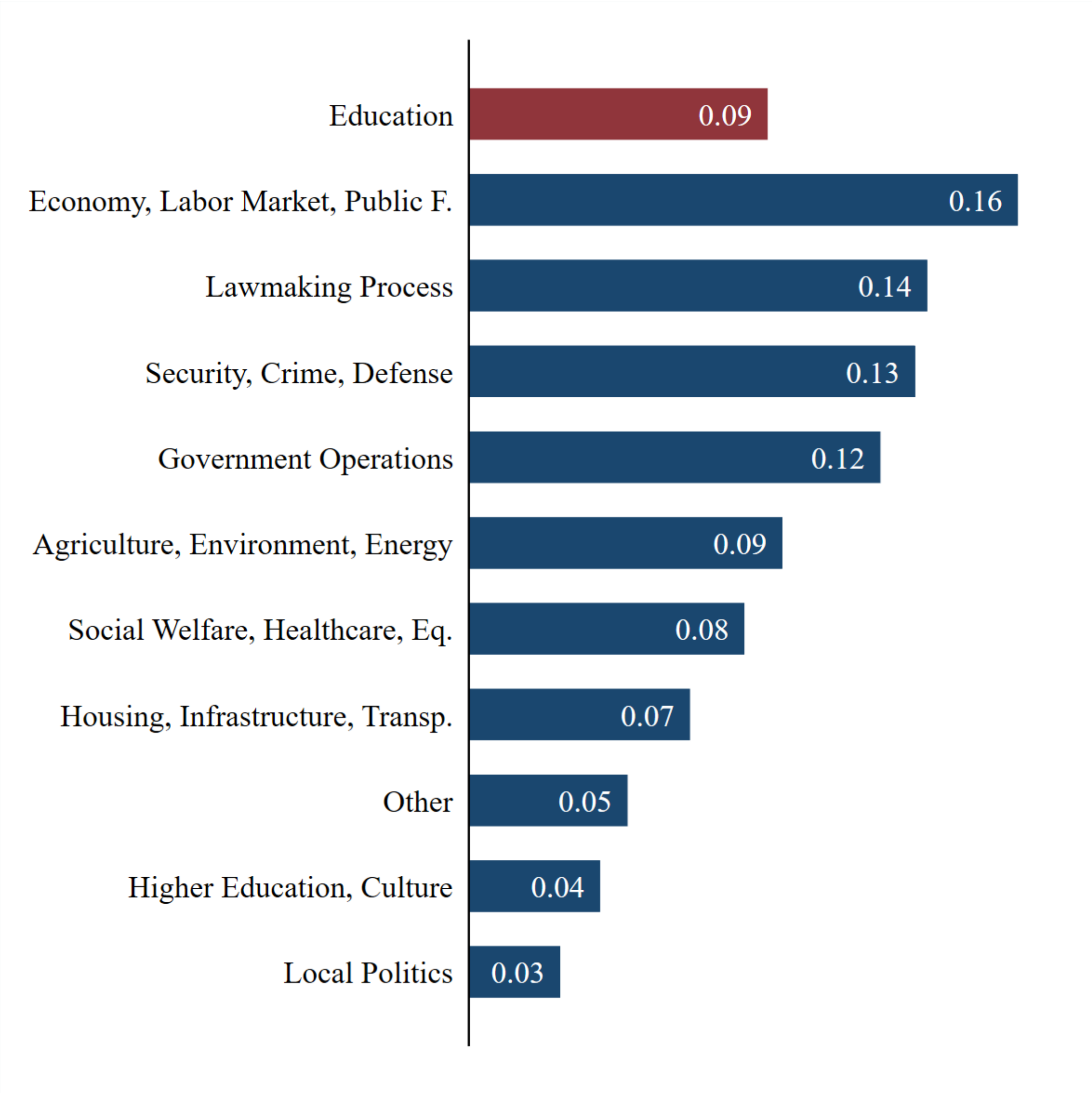


FIGURE A1: SHARE OF SPEECHES' TOPICS

Notes: I classified the topic "education", in red, with a supervised machine learning algorithm. The remaining topics have been classified using an unsupervised machine learning algorithm, namely correlated topic modelling, and assigning the topic with the highest weight to each speech. Details of the classification task are provided in Appendix C.

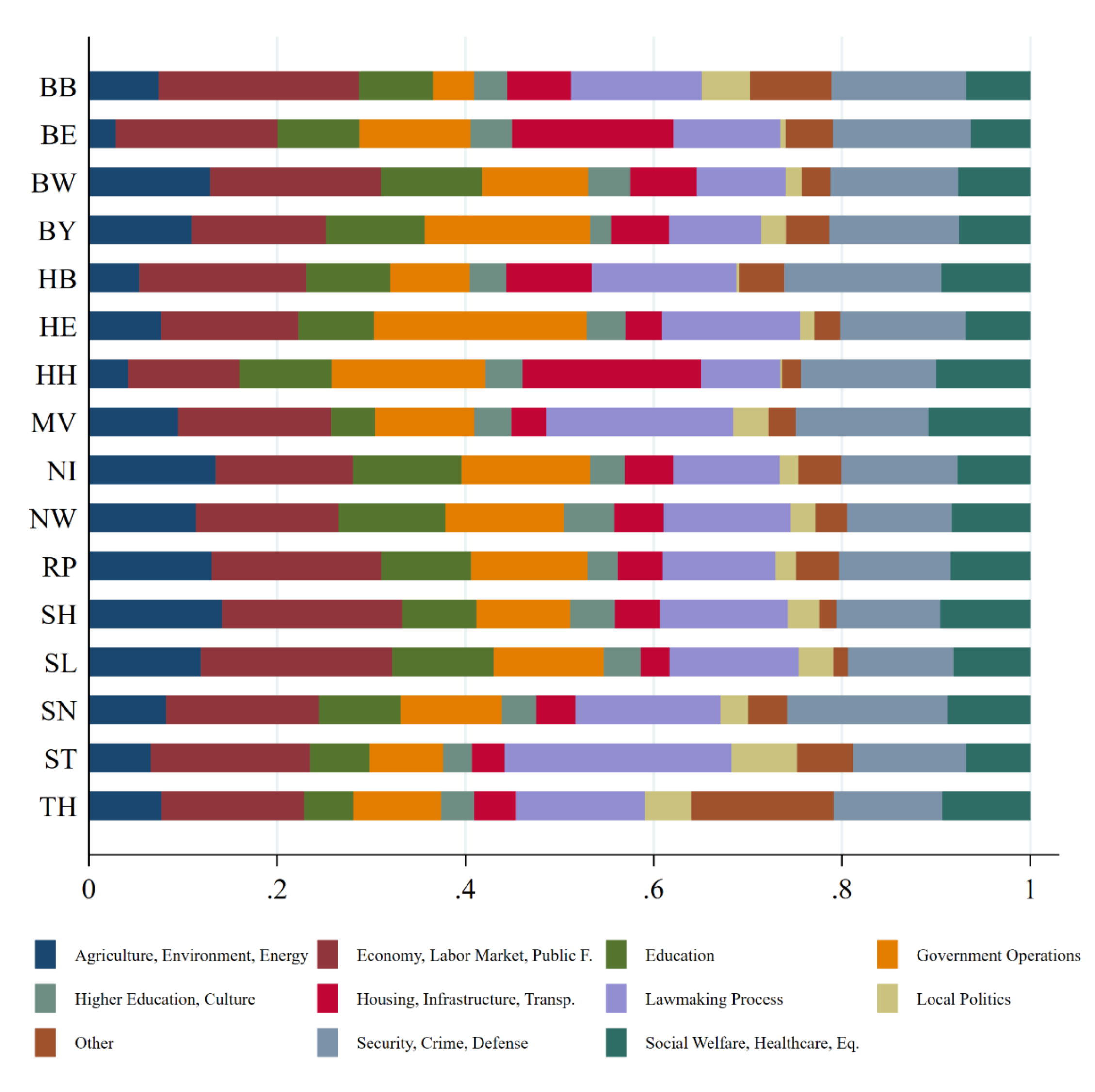


FIGURE A2: SHARE OF SPEECHES' TOPICS, BY STATE

Note: Share of speeches' topics by state. Details of the topic classification task are provided in Appendix C. The codes identifying German states on the y-axes are the official 2-letter acronyms and correspond to the following states: Brandenburg (BB), Berlin (BE), Baden-Württemberg (BW), Bavaria (BY), Bremen (HB), Hessen (HE), Hamburg (HH), Mecklenburg-Vorpommern (MV), Lower Saxony (NI), North Rhine-Westphalia (NW), Rhineland-Palatinate (RP), Schleswig-Holstein (SH), Saarland (SL), Saxony (SN), Saxony-Anhalt (ST), Thuringia (TH).

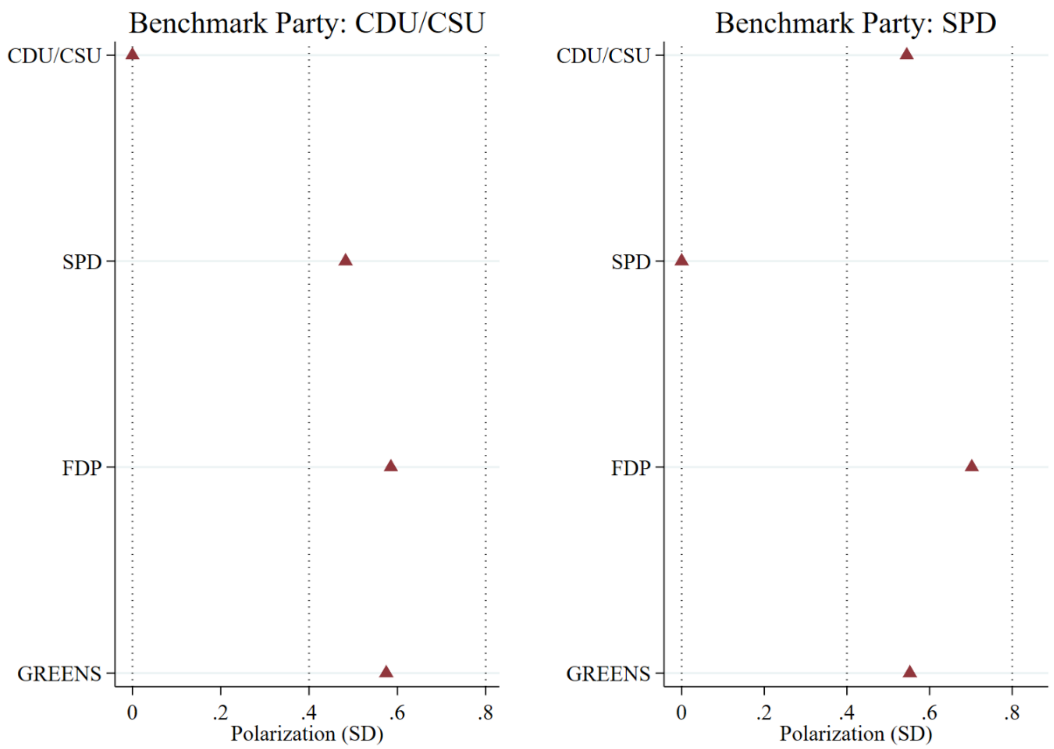


FIGURE A3: POLARIZATION BY PARTY

Notes: The figure reports the average polarization measure aggregated at the party level with respect to a benchmark party. The polarization measured has been divided by its standard deviation and recentered around the average polarization of the CDU/CSU party on the left panel and around the average polarization of the SPD party on the right panel. The x-axis can therefore be interpreted in terms of standard deviation. The polarization measure consists of the opposite of the cosine similarity between all the speeches from a benchmark party (CDU/CSU in the left panel and SPD in the right panel) and all the other speeches in the same topic and legislative period.

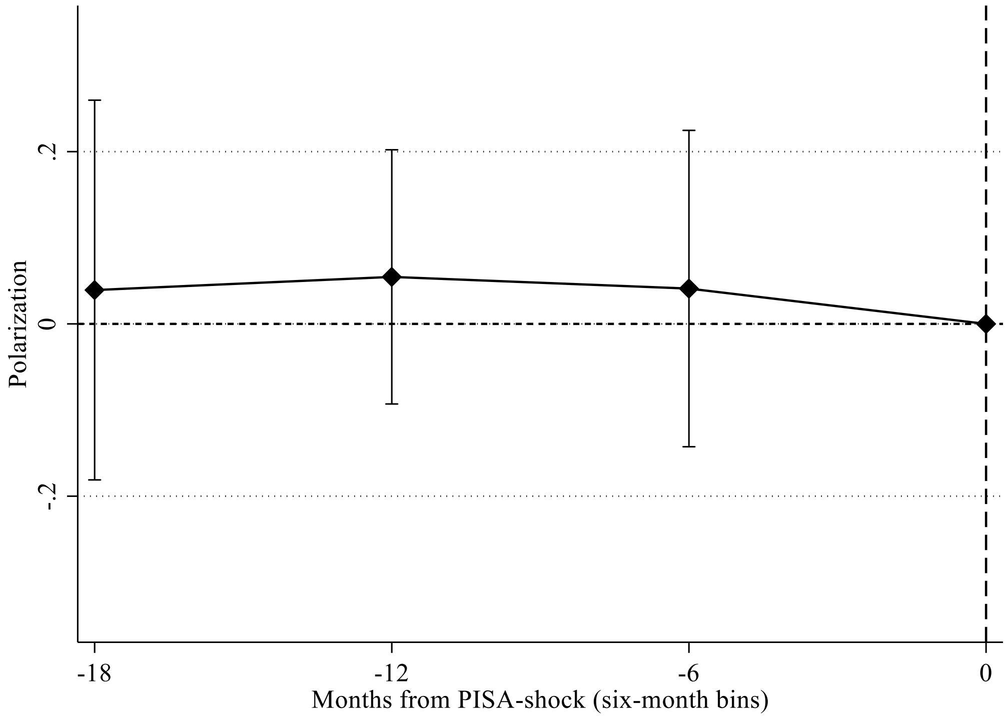


FIGURE A4: PRE-TRENDS IN POLARIZATION

Notes: The graph plot coefficients and 95% confidence interval from the interaction between the dummy variable indicating whether a speech is about education and six-month fixed effects. Standard errors have been clustered at the state level. The dependent variable is the standardized polarization. Only pre-trends are reported. The x-axis reports the distance (in six-month bins) from the PISA shock, which occurred on the 4th of December 2001. The six-month bin prior to the PISA shock is the excluded category. Standard errors are clustered at the state level. The *p*-values of the joint hypothesis test of the pre-trend coefficients being different from 0 is .901.

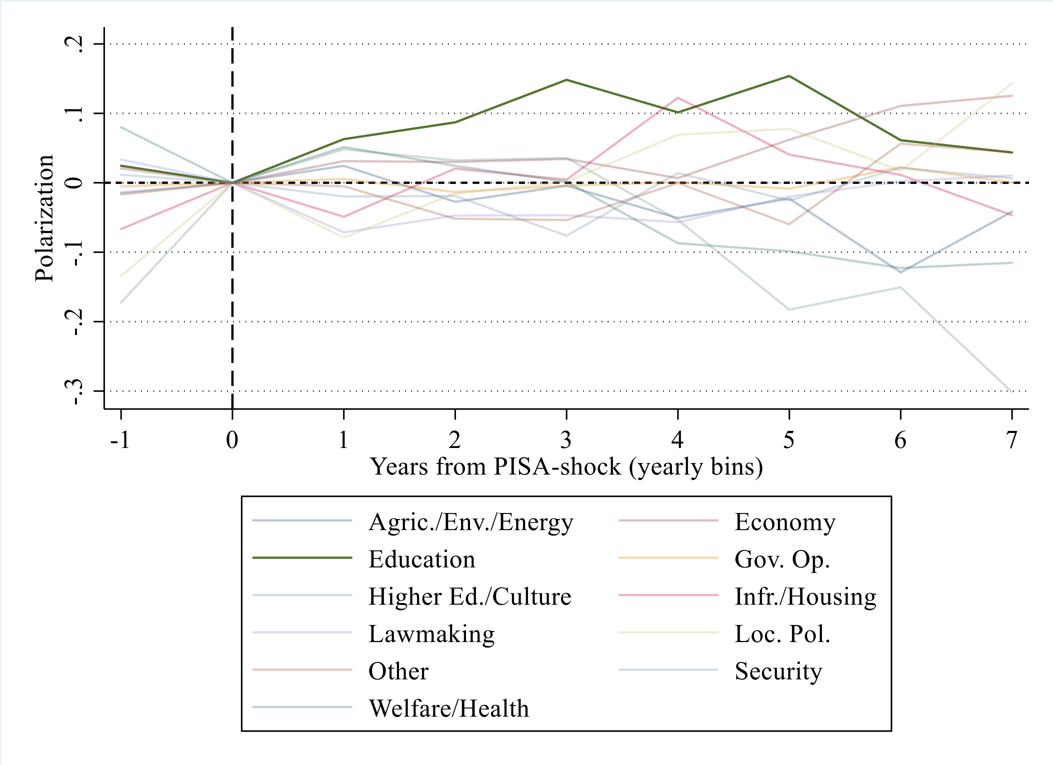


FIGURE A5: TRENDS IN RESIDUALIZED POLARIZATION BY TOPIC

Notes: The figure reports the average standardized and residualized polarization measure over time for each topic. The polarization measure has been residualized of the controls and fixed effects in Equation (4). The measure has been normalized to 0 in the year before the shock for each topic.

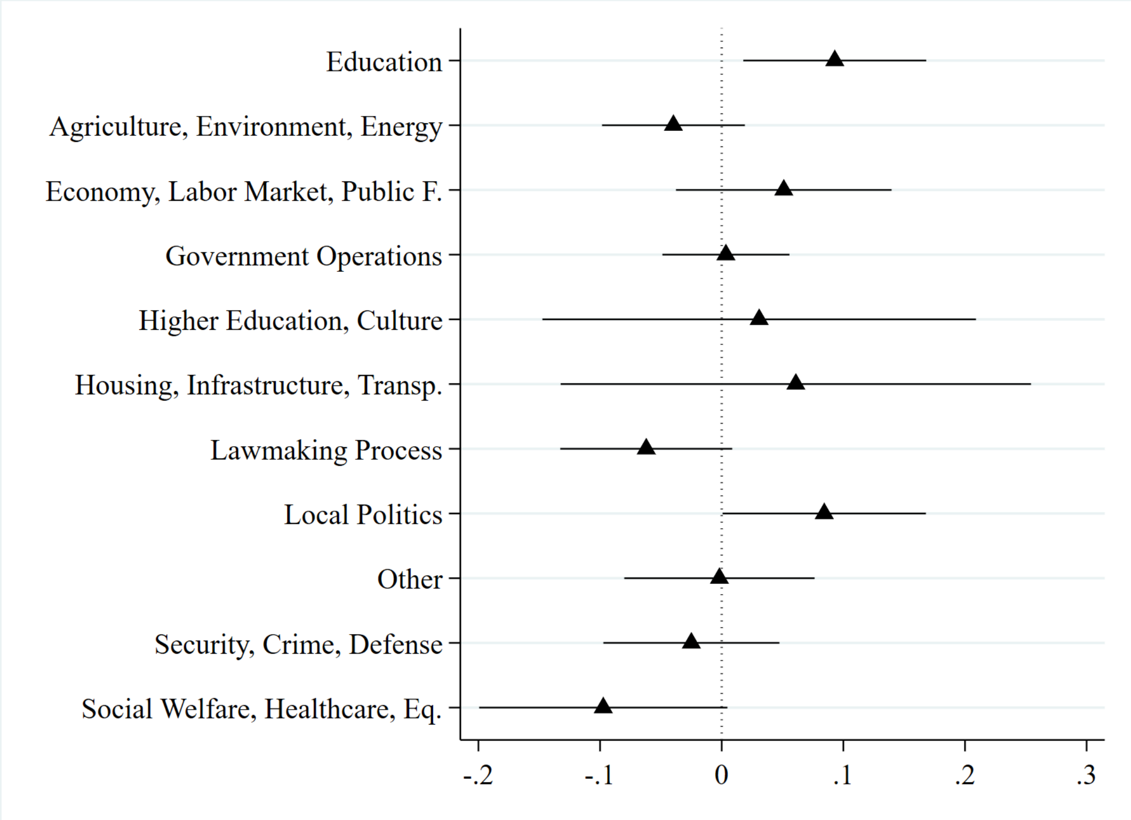


FIGURE A6: PLACEBO WITH OTHER TOPICS

Notes: Difference-in-differences estimate of the impact of the PISA shock on the polarization of the topic indicated in each row with 95% confidence intervals. The dependent variable is the standardized polarization with CDU/CSU as the benchmark party. All regressions include a dummy for whether the speeches occurred after the PISA shock, topic, state-legislative period, party, and year fixed effects. Controls include the length of a speech, the shares obtained at the latest state election by the two main parties, CDU/CSU and SPD, and a dummy variable for whether a speech is given by a minister or a state secretary, and if the MPs belongs to the governing coalition, and the distance from the next election. The data include all parliamentary debates from January 2000 till August 2008. Standard errors have been clustered at the state level. In the first row, I report the coefficient of the impact of the PISA shock on the polarization of education speeches, i.e, the "true" shock.

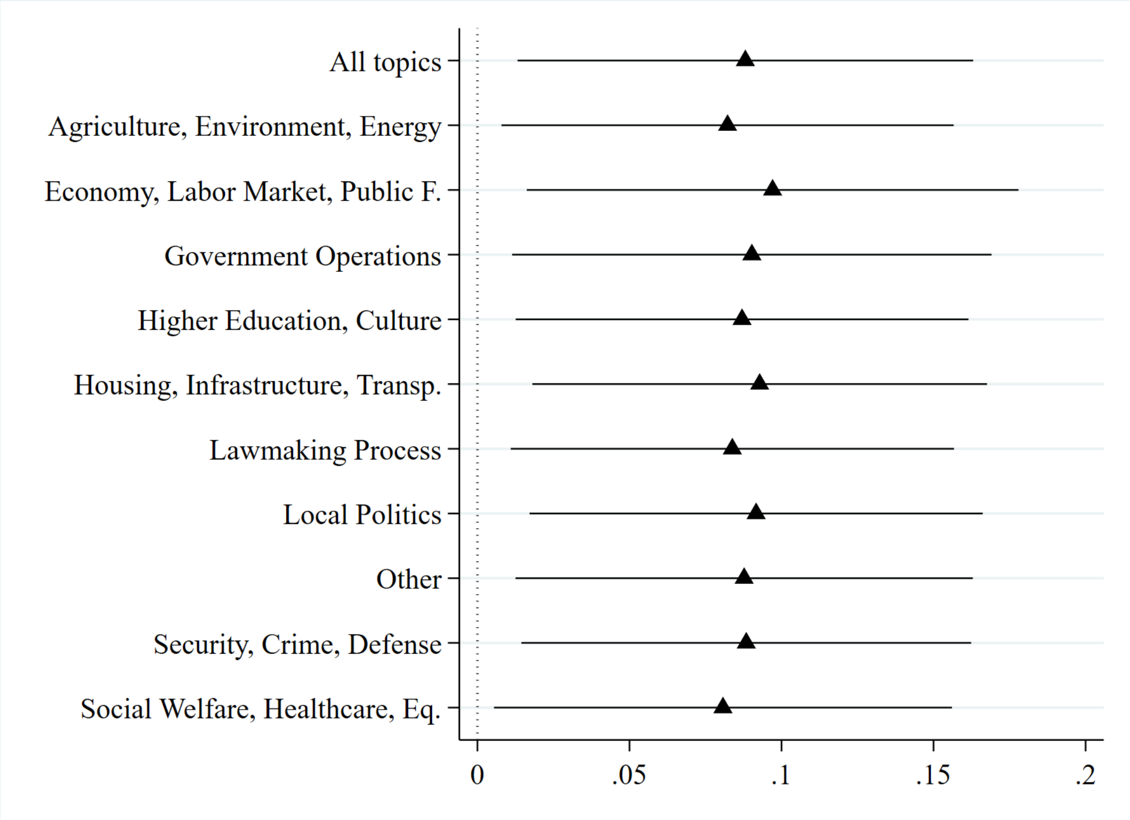


FIGURE A7: LEAVE-ONE-TOPIC-OUT

Notes: Difference-in-differences estimate of the impact of the PISA shock on the polarization of the education topic obtained by dropping the topic indicated in each row with 95% confidence intervals. In the first row, all topics are included. The dependent variable is the standardized polarization with CDU/CSU as the benchmark party. All regressions include a dummy for whether the speeches occurred after the PISA shock, topic, state-legislative period, party, and year fixed effects. Controls include the length of a speech, the shares obtained at the latest state election by the two main parties, CDU/CSU and SPD, and a dummy variable for whether a speech is given by a minister or a state secretary, and if the MPs belongs to the governing coalition, and the distance from the next election. The data include all parliamentary debates from January 2000 till August 2008. Standard errors have been clustered at the state level.

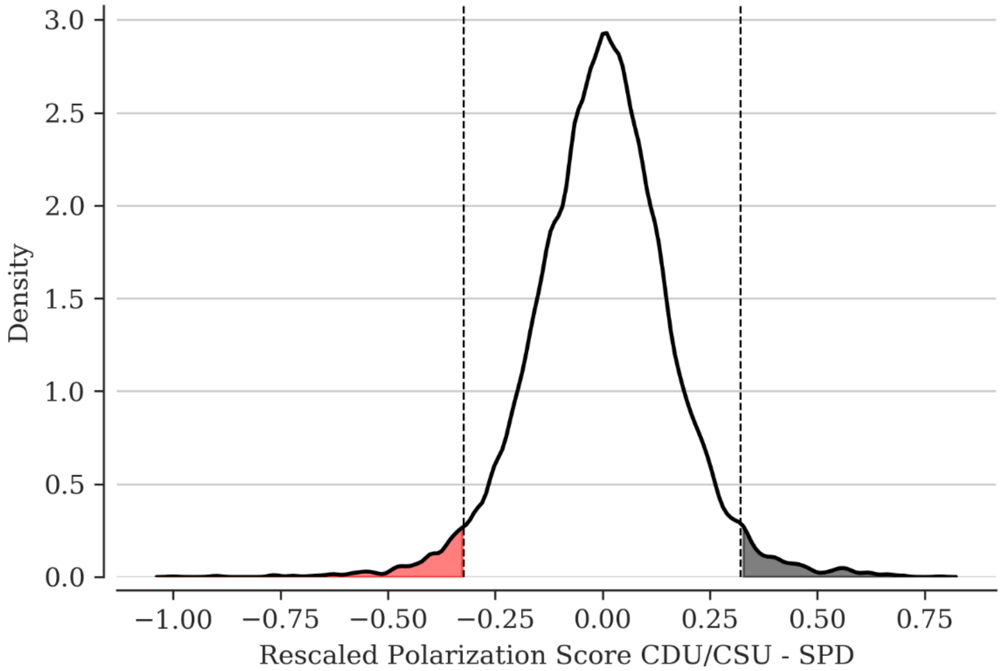
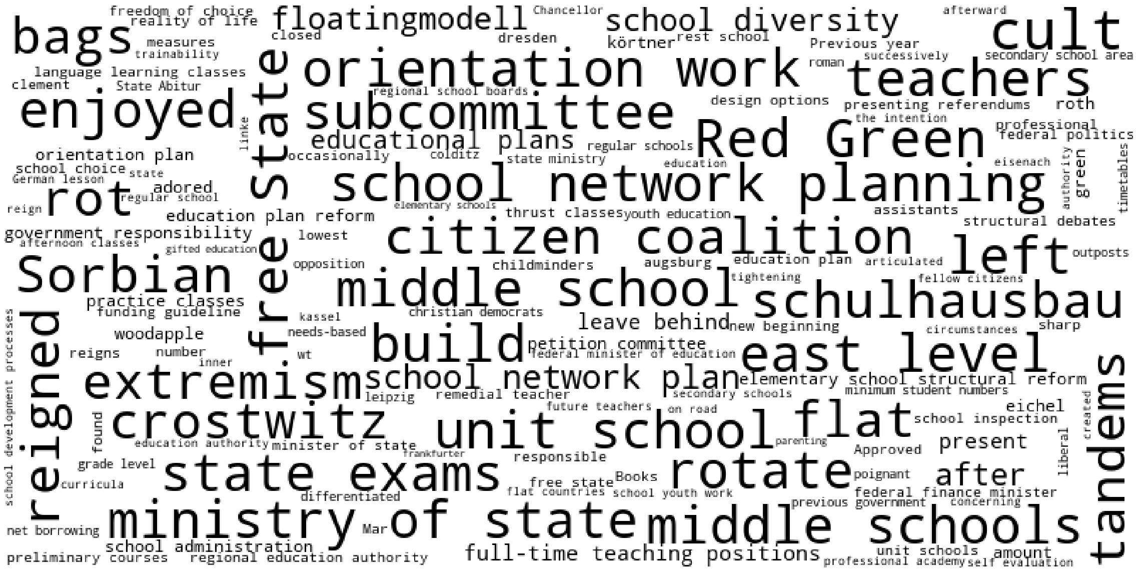


FIGURE A8: DENSITY PLOT OF RESCALED POLARIZATION SCORE (CDU/CSU – SPD)

Notes: The figure reports the density plot of the rescaled polarization score between the CDU/CSU and the SPD. Positive (negative) values indicate terms that are uttered more often by MPs of the CDU/CSU (SPD) with respect to members of the SPD (CDU/CSU). The score has been centered around 0 and divided by the maximum of the absolute value of the score max $(|p(w)|)$, so that $-1 \leq p(w) \leq 1$. Terms with polarization scores in the black-(red-)shaded area are the top-250 terms in terms of polarization score for the CDU/CSU (SPD) and are depicted in Figure A9 (Figure A10).

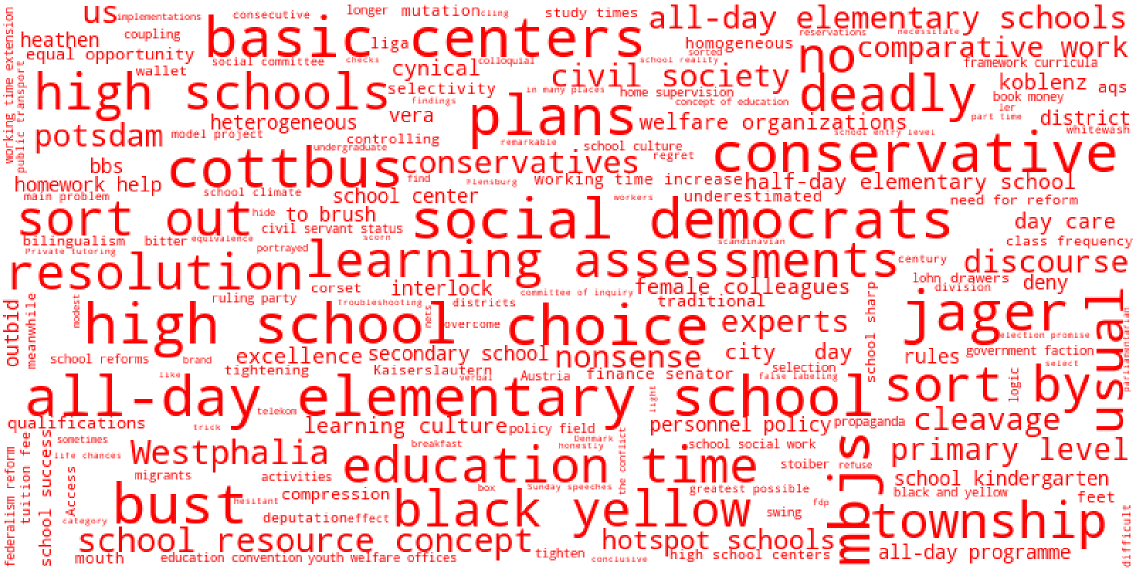


(a): English (translated)

(b): German (original)

FIGURE A9: MOST POLARIZING WORDS - CDU

Notes: The figure reports the 250 most polarizing words for the CDU/CSU. In Panel (a), the words have been translated into English using the Python package *deep_translator.* In Panel (b), I report the original German terms. The font size of the words increases with the polarization score $p(w)$ of each term.



(a): English (translated)

(b): German (original)

FIGURE A10: MOST POLARIZING WORDS - SPD

Notes: The figure reports the 250 most polarizing words for the SPD. In Panel (a), the words have been translated into English using the Python package *deep_translator*. In Panel (b), I report the original German terms. The font size of the words increases with the polarization score $p(w)$ of each term.

TABLE A1—STATE-SPECIFIC RESULTS IN PISA 2000

State	PISA State Score Reading (1)	Deviation from federal mean (2)	Position in international PISA ranking (3)
Bavaria	510	26	11
Baden-Württemberg	500	16	18
Saxony	491	7	23
Rhineland-Pfalz	485	1	25
Saarland	484	0	27
North Rhine-Westphalia	482	-2	29
Thuringia	482	-2	30
Schleswig-Holstein	478	-6	33
Hessen	476	-8	34
Lower Saxony	474	-10	36
Mecklenburg-Vorpommern	467	-17	38
Brandenburg	459	-25	40
Saxony-Anhalt	455	-29	42
Bremen	448	-36	44

Notes: The table reports the average performance in reading of each German state in Column 1, the distance from the average German performance in Column 2, and position in the international PISA ranking in Column 3. Data have been taken from OECD (2001). Results for Berlin and Hamburg were not made public due to these states not meeting the prescribed threshold of sample size.

TABLE A2— BILLS BY TOPIC AND STATUS

Topic	Number of Bills	Share	Topic	Number of Bills	Share
Education	525	0.10	Taxes & Dues	67	0.01
Political System & Parties	507	0.09	Social Matters	67	0.01
Other	426	0.08	Justice and Laws	51	0.01
Communal Matters	376	0.07	Social Welfare	49	0.01
State Budget	269	0.05	Housing	44	0.01
Justice and Security	232	0.04	Europe	43	0.01
Government Officials	230	0.04	Regional Planning	41	0.01
Health	228	0.04	Immigration & Integration	40	0.01
Economy	223	0.04	Lottery/Gambling Industry	40	0.01
Environment	185	0.03	Culture	35	0.01
Labor	177	0.03	Animals	33	0.01
Media	175	0.03	Data Protection	29	0.01
Family/Children/Youth	164	0.03	Civic Duties	29	0.01
Administration	140	0.03	Data	28	0.01
Taxes & Finances	138	0.03	Pension/Seniority/Retirement Planning	27	0.01
Judicial System	112	0.02	Agriculture	25	0
Construction	110	0.02	Religion	23	0
Finances	99	0.02	Technology	19	0
Equality	89	0.02	Energy	15	0
Civil rights	83	0.02	Community Financing	11	0
Traffic and Transportation Systems	75	0.01	Defense	6	0
Society	70	0.01	International Matters	1	0
Total Number of Bills			5,356		
Total Number of Accepted Bills			4,116		
Total Number of Rejected Bills			821		
Total Number of Bills with “Other” Status			419		

Notes: The table reports the total number and share of bills by topic, and the total number of bills by status (accepted, rejected, or with “other” status) for the period January 2000 – August 2008. The data come from the “Patterns of Lawmaking in the German Lander” dataset (Stecker, Kachel, and Paasch 2021). The original topic names in German can be found in Stecker, Kachel, and Paasch (2021).

TABLE A3—THE EFFECT OF THE PISA SHOCK ON THE SHARE OF EDUCATION SPEECHES

	(1)	(2)	(3)	(4)	(5)
PISA shock (*Federal*)	0.018***	0.024***	0.024***	0.021***	0.022***
	(0.004)	*(0.006)*	*(0.005)*	*(0.006)*	*(0.006)*
PISA shock (*State*)		-0.008*	0.004	-0.009	-0.117
		(0.004)	*(0.008)*	*(0.009)*	*(0.110)*
PISA shock *(State)* × PISA score (Dummy)			-0.014*		
			(0.007)		
PISA shock (*State*) × Med. Perf.				-0.002	
				(0.009)	
PISA shock (*State*) × High Perf.				0.009	
				(0.009)	
PISA shock (*State*) × PISA Perf./100					0.023
					(0.023)
Mean DV (Pre-shock)			0.073		
State FE	Yes	Yes	Yes	Yes	Yes
Party FE	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes
R^2	0.009	0.009	0.009	0.009	0.009
Observations	210,006	210,006	210,006	185,729	185,729

Notes: The table reports OLS estimate of the impact of the PISA shock on the share of education speeches. The dependent variable is a dummy variable indicating whether a speech is about education. The variable, "PISA shock (*Federal*)" is a dummy variable which takes value one if a speech occurred after 4th December 2001, when the first PISA results were released. The variable "PISA shock (*State*)" is a dummy variable that takes value one if a speech occurred after 26th June 2002, the date on which state specific results were released. The PISA-score-dummy variable takes value one if the state-specific results of PISA 2000 were released for the corresponding states, i.e., it takes value 0 for the states of Berlin and Hamburg and 1 for the other states. The medium performance variable takes value one if the performance of the respective state is in the middle tercile, while high performance takes value one if the performance is in the upper tercile. The "PISA Performance" variable represents the performance on each state in the reading test as reported in Table A1 (column 1). Controls include the length of a speech, the shares obtained at the latest state election by the two main parties, CDU/CSU and SPD, and a dummy variable for whether a speech is given by a minister or a state secretary, and if the MPs belongs to the governing coalition, and the distance from the next election.. The data include all parliamentary debates from January 2000 till August 2008. Standard errors (in parentheses) have been clustered at the state level. *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

TABLE A4—HETEROGENEITIES BY PARTY – LEFT-RIGHT POLARIZATION MEASURE

	(1)	(2)	(3)	(4)	(5)
PISA shock × Education	0.068	0.078**	0.090**	0.093**	0.012
	(0.039)	*(0.031)*	*(0.033)*	*(0.036)*	*(0.056)*
PISA shock × Education × CDU/CSU	0.047				0.103**
	(0.030)				*(0.048)*
PISA shock × Education × SPD		0.026			0.092*
		(0.038)			*(0.045)*
PISA shock × Education × FDP			-0.038		0.040
			(0.041)		*(0.060)*
PISA shock × Education × GREENS				-0.053	0.028
				(0.057)	*(0.068)*
Topics, State-Legisl. Period, Party, Year FE	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes
R^2	0.553	0.553	0.553	0.553	0.553
Observations	205,160	205,160	205,160	205,160	205,160

Notes: Difference-in-differences estimate of the impact of the PISA shock on the polarization of education speeches. The dependent variable is the standardized cosine similarity between each speech from the left (right) parties and all the speeches from the right (left) in the same topic, state and legislative period. All regressions include a dummy for whether the speeches occurred after the PISA shock, topic, state-legislative period, party, and year fixed effects. Controls include the length of a speech, the shares obtained at the latest state election by the two main parties, CDU/CSU and SPD, and a dummy variable for whether a speech is given by a minister or a state secretary, and if the MPs belongs to the governing coalition, and the distance from the next election. The data include all parliamentary debates from January 2000 till August 2008. Standard errors (in parentheses) have been clustered at the state level. *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

TABLE A5—SYMMETRIC TIME WINDOW (2000-2004)

	(1)	(2)	(3)
PISA shock × Education	0.076*	0.101**	0.096**
	(0.041)	*(0.036)*	*(0.036)*
Topic FE	Yes	Yes	Yes
State-Legislative Period FE	No	Yes	Yes
Party, Year FE	No	Yes	Yes
Controls	No	No	Yes
R^2	0.139	0.263	0.550
Observations	72,784	72,784	72,784

Notes: Difference-in-differences estimate of the impact of the PISA shock on the polarization of education speeches. The dependent variable is the standardized polarization with CDU/CSU as the benchmark party. All regressions include a dummy for whether the speeches occurred after the PISA shock and topic fixed effects. Controls include the length of a speech, the shares obtained at the latest state election by the two main parties, CDU/CSU and SPD, and a dummy variable for whether a speech is given by a minister or a state secretary, and if the MPs belongs to the governing coalition, and the distance from the next election. The data include all parliamentary debates from January 2000 till December 2004. Standard errors (in parentheses) have been clustered at the state level. *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

TABLE A6—DIFFERENT NUMBER OF TOPICS

	Baseline	Number of Topics				
		9	10	11	13	15
	(1)	(2)	(3)	(4)	(5)	(6)
PISA shock × Education	0.088**	0.090**	0.077**	0.083**	0.071*	0.063*
	(0.038)	*(0.037)*	*(0.036)*	*(0.037)*	*(0.035)*	*(0.035)*
Topic, State-Legisl. Period, Party, Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.535	0.559	0.561	0.555	0.553	0.546
Observations	137,820	138,274	138,124	137,850	137,427	137,238

Notes: Difference-in-differences estimate of the impact of the PISA shock on the polarization of education speeches. In column 1, I report the baseline estimate using the CTM estimated with 30 topics, aggregated into 11 topics. In column 2, 3, 4, 5 and 6, I report the estimate using the CTM estimated using 9, 10, 11, 13, and 15 topics, respectively. The dependent variable is the standardized polarization with CDU/CSU as the benchmark party. All regressions include a dummy for whether the speeches occurred after the PISA shock and topic fixed effects. Controls include the length of a speech, the shares obtained at the latest state election by the two main parties, CDU/CSU and SPD, and a dummy variable for whether a speech is given by a minister or a state secretary, and if the MPs belongs to the governing coalition, and the distance from the next election. The data include all parliamentary debates from January 2000 till August 2008. Standard errors (in parentheses) have been clustered at the state level. *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

TABLE A7—SYMMETRIC TIME WINDOW (2000-2004) WITH EDUCATION AND PLACEBO TOPICS (LOCAL POLITICS AND SOCIAL WELFARE/HEALTHCARE)

	Education	Local Politics	Social Welfare/Healthcare
	(1)	(2)	(3)
PISA shock × Topic Dummy	0.096**	0.042	-0.014
	(0.036)	*(0.064)*	*(0.041)*
Topic, State-Legisl. Period, Party, Year FE	Yes	Yes	Yes
Controls	Yes	Yes	Yes
R^2	72,784	72,784	72,784
Observations	137,820	137,820	137,820

Notes: Difference-in-differences estimate of the impact of the PISA shock on the polarization of speeches about education (column 1), about local politics (column 2), and about social welfare/healthcare (column 3). The dependent variable is the standardized polarization with CDU/CSU as the benchmark party. All regressions include a dummy for whether the speeches occurred after the PISA shock and topic fixed effects. Controls include the length of a speech, the shares obtained at the latest state election by the two main parties, CDU/CSU and SPD, and a dummy variable for whether a speech is given by a minister or a state secretary, and if the MPs belongs to the governing coalition, and the distance from the next election. The data include all parliamentary debates from January 2000 till December 2004. Standard errors (in parentheses) have been clustered at the state level. *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

TABLE A8—SENSITIVITY TO DIFFERENT THRESHOLDS OF TERM FREQUENCY

	Lower Bound of Term Frequency					
	> 20 Speeches (1)	> 30 Speeches (2)	> 40 Speeches (3)	> 2% Speeches (4)	> 2.5% Speeches (5)	> 5% Speeches (6)
PISA shock × Education	0.088**	0.094**	0.097**	0.083**	0.080**	0.063**
	(0.039)	(0.038)	(0.037)	(0.036)	(0.037)	(0.029)
Topic, State-Legisl. Period, Party, Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.536	0.537	0.537	0.470	0.459	0.417
Observations	137,820	137,820	137,820	137,820	137,820	137,820

Notes: Difference-in-differences estimate of the impact of the PISA shock on the polarization of education speeches. The dependent variable is the standardized polarization with CDU/CSU as the benchmark party computed using only words that appear in at least 20 speeches in a topic (column 1), at least 30 speeches (column 2), 40 (column 3), 2% (column 4), 2.5% (column 5) and 5% (column 6). All regressions include a dummy for whether the speeches occurred after the PISA shock, topic, state-legislative period, party, and year fixed effects. Controls include the length of a speech, the shares obtained at the latest state election by the two main parties, CDU/CSU and SPD, and a dummy variable for whether a speech is given by a minister or a state secretary, and if the MPs belongs to the governing coalition, and the distance from the next election. The data include all parliamentary debates from January 2000 till August 2008. Standard errors (in parentheses) have been clustered at the state level. *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

Appendix B – Corpus Collection

The main data source for this project consists of parliamentary debates from the 16 German state parliaments for the period 2000-2008. The transcripts of such debates are not available in a structured format. They can be retrieved from each state-parliaments' website as PDF documents.[34] I have obtained these documents by web scraping each state-parliaments' website. Each document contains the transcript of a single plenary debate. Most transcripts of debates held in the period 2000-2008 are available as machine-readable PDF documents (93%). The remaining transcripts (7%) need to be first converted into machine-encoded text through an Optical Character Recognition (OCR) software. The share of documents for which OCR is necessary increases dramatically for debates held before 2000. This step is error-prone and renders the parsing of documents less reliable. Further, the availability of these documents on state-parliaments' websites decreases for debates held before 2000. For these reasons, I limit my analysis to debates from January 2000. In total, I have collected 3,302 PDF documents, 206.4 per state on average. In Figure B1, I report an example of a page from a plenary debate in the state of Baden-Württemberg that occurred on the 13th of December 2001. The raw text is clearly readable, but it lacks a formal structure. For my analysis, it is necessary to systematically identify and process the different features of the document, such as the name and role of the speaker, the party to which she belongs, the speech, the interruptions, the header, the page number etc.

In the example, the first speaker is Ms. Renate Rastätter, written in bold, a member of the parliament, as denoted by the abbreviation *Abg.* (*Abgeordnete*, member of parliament in German), of the Green Party (*GRÜNE*). The speech starts directly thereafter, and interruptions are reported in parentheses and are indented. At the end of the page, the President of the session and a Minister also speak. It can be noted how the way speaker names are reported also depends on their role.

These features constitute only some of the challenges that need to be addressed. The process to transform the PDF documents into a structured dataset suitable for my analysis involves four main steps, which I now briefly describe.[35] Each step described has to be run separately for each state, as the process needs to be adapted to the different

[34] The only exception is Saarland, which does not provide the transcript of the parliamentary debates for the period 2000-2008 on its website. These debates were made available for my research upon request to the administration of the parliament.

[35] The process builds upon the publicly available GitHub repository https://github.com/panoptikum/plenary_record_parser , which contains the codes used by Felix Idelberger to perform the same task for German state parliamentary debates for the period 2008-2018.

structure of the PDF documents in each state. For example, the way speakers and interruptions are reported in the transcripts differ substantially across states.

B1. – Layout scan.

The aim of the first step of the pipeline is to identify the coordinates that identify the location of the different elements in the document. This allows me to process the different features of the text correctly. To this purpose, I scan the layout of all the PDFs in a state using the Python package *layout_collector*. During this process, the coordinates of each text box that contains any content in each page of the documents are recorded. After all the coordinates have been recorded, I analyze the distribution of the coordinates of the text boxes to infer the relevant coordinates for the main text and interruptions in the left and right column, the header and the footer. I record such coordinates, which will be used in the next steps.

B2. – Conversion into XML files

All the PDFs are converted into XML files using the Python package *pdf2txt*. The XML version of each PDF stores information for each element in the PDF files, such as position, font, font size, boldness etc.

B3. – Conversion into plain text files

In this step, I convert the XML documents into plain text files enriched by features of the original PDF document recorded in the XML version. I use the coordinates for the headers and page numbers recorded in step B1 to drop the headers and page numbers while reconstructing the text files. Similarly, I use the coordinates recorded for the main text to ensure that the text on the left column of each page precedes the text on the right column. I also insert tags to denote interruptions and words in bold. Figure B2 shows the outcome of this step for the page depicted in Figure B1. One can observe that the plain text file does not contain headers nor page numbers, and that the text in the left column precedes the text in the right column. The text files are also enriched with tags to identify interruptions and speakers. These are added every time the text is indented with respect to the column in which it is located or when it is written in bold in the original PDF document (*<indentation>* and *<poi_begin>* followed by *<poi_end>*), respectively.

B4. – Parsing

In the last step, I process each plain text file line-by-line. The processing script contains an exhaustive collection of regular expressions which capture the features of the documents, like the name of the speaker, party, role etc. and processes them accordingly. The script collects the speeches of each speaker as well as interruptions. Each speech or interruption is assigned to the speaker who utters them or, in the case of interruptions, under which they occur. Together with the name of the speaker, a variety of metadata such as the party (if reported), role, date, state, legislative session etc. are also collected. All the speeches are then aggregated into a single dataset where each observation corresponds to either a speech or an interruption and all the corresponding metadata. For the purpose of my analysis, I drop all the interruptions and aggregate speeches as a single utterance. In Figure B2 below, for example, an observation would correspond to the entire speech by Ms. Renate Rastätter until the speech of the President, stripped of all the interruptions.

B5. – Aggregation

Finally, the processing script aggregates all the speeches into a unified corpus. The process occurs hierarchically. At the end of each plenary debate, all the speeches are stored into a dataset. Once all the plenary debates in a state have been processed, all the separate debates are aggregated into a single dataset, which consists of all the speeches uttered in the period 2000-2008 in a single state and the related metadata (speaker, party affiliation, role, date, state etc.). As a final step, all the speeches of all 16 states in the period 2000-2008 are aggregated into one corpus, which provides the main data source for this project.

Landtag von Baden-Württemberg – 13. Wahlperiode – 15. Sitzung – Donnerstag, 13. Dezember 2001

Abg. Renate Rastätter GRÜNE: Herr Präsident, meine Damen und Herren! Die Ergebnisse der PISA-Studie sind in der Tat Grund zur Sorge und Anlass zum Handeln. Nicht hilfreich ist es allerdings, in Panik und Angst zu verfallen, wie das jetzt bei vielen der Fall ist. Zum Beispiel fordert der CDU-Politiker Rüttgers jetzt einen Sprachtest für alle Dreijährigen. Oder ich nenne die Aussage des Rektors eines Gymnasiums in Nordbaden, der jetzt fordert: Wir brauchen mehr Pauken von Faktenwissen statt Orientierungswissen. Oder es gibt die Äußerung von Ministerpräsident Stoiber, der seine politische Forderung nach einem Einwanderungsstopp jetzt mit den schlechten Schulleistungen von ausländischen Kindern begründet. Das, meine Damen und Herren, betrachte ich als eine infame Instrumentalisierung

(Zuruf von der CDU)

der PISA-Studie für politische Zwecke.

(Beifall bei den Grünen und Abgeordneten der SPD)

Diese Äußerung wird auch wider besseres Wissen gemacht. Denn wenn die PISA-Studie eines gezeigt hat, dann dies, dass es in der Bundesrepublik Deutschland nicht gelingt, gerade ausländische Schüler in den vorschulischen Einrichtungen und in der Grundschule ausreichend zu fördern, sodass sie in unserem System bessere Bildungschancen haben.

(Zuruf von der CDU)

Auch Sie, Frau Ministerin Schavan, wissen sofort, worin die Ursachen der Misere liegen. Sie zählen wieder alle Rezepte auf, alle Ihre Reformprojekte: frühere Einschulung, achtjähriges Gymnasium, reformierte Oberstufe, Englisch an Grundschulen, Neuorientierung des Unterrichts. Gleichzeitig pflegen Sie die altbekannten Vorurteile: Gesamtschulen seien schlecht, „Kuschelecken" ersetzten nicht das nachhaltige Lernen.

(Beifall bei Abgeordneten der CDU – Zuruf des Abg. Oelmayer GRÜNE)

Das, meine Damen und Herren, halte ich für ein populistisches Ausspielen

(Unruhe)

von Leistung und Sich-wohl-Fühlen. Es ist eine Diskriminierung von Grundschullehrerinnen in unserem Land. Diese nehmen nämlich den Erziehungs- und Bildungsauftrag des Lehrplans der Grundschule ernst, der fordert, Grundschulen als Lern- und Lebensorte auszugestalten, als Orte, an denen sich Kinder auch wohl fühlen können.

(Beifall bei den Grünen und des Abg. Zeller SPD)

Dabei gibt es durchaus Reformprojekte, die wir unterstützen, Frau Ministerin: zum Beispiel die frühere Einschulung gekoppelt mit dem „Schulanfang auf neuen Wegen", die jahrgangsübergreifenden Klassen, bei denen das Prinzip gilt: differenzieren und fördern.

Wenn man sich allerdings anschaut, wie die Entwicklung verläuft, muss man sagen: Nachdem Sie sich bundesweit damit profiliert haben, haben Sie das Interesse verloren. Das Reformprojekt dümpelt vor sich hin. Stattdessen hätte es ein Schlüsselprojekt für die Weiterentwicklung der Grundschule werden können. Nur 4 % der 2 500 Grundschulen haben dieses Projekt tatsächlich durchgeführt.

Ich vermisse somit die Bereitschaft, Frau Kultusministerin, innezuhalten und auch einmal kritisch zu fragen: Mache ich, machen wir in diesem Bundesland eigentlich alles richtig? Man darf nicht immer nur sagen: Wir sind Spitze, wir können alles.

(Beifall der Abg. Heike Dederer GRÜNE – Abg. Pfister FDP/DVP: Außer Hochdeutsch!)

Nachdenklichkeit, meine Damen und Herren, ist allerdings auch in diesem Hause, bei uns selbst, angesagt. Ich würde es für ein gutes Zeichen halten, wenn sich der Landtag entschließen könnte, eine Enquetekommission zum Thema „Weiterentwicklung von Schule und Unterricht" einzusetzen. Ich möchte daran erinnern – Frau Kollegin Rudolf hat dies ja bereits angesprochen –: Wir sind sehr zufrieden damit, dass wir in der letzten Legislaturperiode die Enquetekommission „Jugend – Arbeit – Zukunft" hatten. Sie hat genau zu der Erkenntnis geführt, dass ungefähr 20 % der Jugendlichen aus sozial benachteiligten Familien nicht mehr die Leistungen erbringen, die für eine berufliche Integration notwendig sind.

Die Jugendenquetekommission hat im Ergebnis ein Bündel von Maßnahmen empfohlen. Unter anderem hat sie die Regierungskoalition endlich davon überzeugt, wie dringend notwendig die Schulsozialarbeit in diesem Land ist. Sie hat dafür gesorgt, dass in Baden-Württemberg Jugendagenturen eingerichtet wurden, die den Jugendlichen helfen, den schwierigen Übergang von der Schule in den Beruf zu meistern.

Deshalb, sage ich, würde es uns gut anstehen, zunächst einmal genau hinzuschauen und zu klären: Wo liegen denn die Schwächen? Vor allem sollten wir aber klären: Welche ganz konkreten Handlungsperspektiven müssen aufgebaut werden, damit alle Jugendlichen, von den sozial benachteiligten bis zu den höchstbegabten, die Bildung bekommen und die Kompetenzen entwickeln können, die sie von ihren Potenzialen her mitbringen?

Ich bedanke mich.

(Beifall bei den Grünen und des Abg. Kaufmann SPD)

Präsident Straub: Das Wort erteile ich Frau Ministerin Schavan.

Ministerin für Kultus, Jugend und Sport Dr. Annette Schavan: Herr Präsident, meine sehr verehrten Damen und Herren! 1997 hat sich die Kultusministerkonferenz in Konstanz entschieden, künftig deutsche Schulen an internationalen Vergleichsstudien zu beteiligen. In den letzten zehn Tagen habe ich mich an diese Situation, an die damalige Sitzung und die Wochen und Monate danach erinnert, und ich habe mich übrigens auch an manche schul- und bildungspolitische Debatte der letzten Jahre in diesem Haus erinnert.

(Abg. Röhm CDU: Jetzt kommts!)

738

FIGURE B1: EXAMPLE OF PDF DOCUMENT

<poi_begin>Abg. Renate Rastätter<poi_end> GRUNE: Herr Präsident, meine Damen und Herren! Die Ergebnisse der PISA-Studie sind in der Tat Grund zur Sorge und Anlass zum Handeln. Nicht hilfreich ist es allerdings, in Panik und Angst zu verfallen, wie das jetzt bei vielen der Fall ist. Zum Beispiel fordert der CDU-Politiker Rüttgers jetzt einen Sprachtest für alle Dreijährigen. Oder ich nenne die Aussage des Rektors ei-nes Gymnasiums in Nordbaden, der jetzt fordert: Wir brau-chen mehr Pauken von Faktenwissen statt Orientierungs-wissen. Oder es gibt die Äußerung von Ministerpräsident Stoiber, der seine politische Forderung nach einem Ein-wanderungsstopp jetzt mit den schlechten Schulleistungen von ausländischen Kindern begründet. Das, meine Damen und Herren, betrachte ich als eine infame Instrumentalisie-rung

<indentation>(Zuruf von der CDU)

der PISA-Studie für politische Zwecke.

<indentation>(Beifall bei den Grünen und Abgeordneten der
<identation>SPD)

Diese Äußerung wird auch wider besseres Wissen ge-macht. Denn wenn die PISA-Studie eines gezeigt hat, dann dies, dass es in der Bundesrepublik Deutschland nicht ge-lingt, gerade ausländische Schüler in den vorschulischen Einrichtungen und in der Grundschule ausreichend zu för-dern, sodass sie in unserem System bessere Bildungschan-cen haben.

<indentation>(Zuruf von der CDU)

Auch Sie, Frau Ministerin Schavan, wissen sofort, worin die Ursachen der Misere liegen. Sie zählen wieder alle Re-zepte auf, alle Ihre Reformprojekte: frühere Einschulung, achtjähriges Gymnasium, reformierte Oberstufe, Englisch an Grundschulen, Neuorientierung des Unterrichts. Gleich-zeitig pflegen Sie die altbekannten Vorurteile: Gesamt-schulen seien schlecht, „Kuschelecken“ ersetzten nicht das nachhaltige Lernen.

<indentation>(Beifall bei Abgeordneten der CDU – Zuruf des
<indentation>Oelmayer GRÜNE)

Das, meine Damen und Herren, halte ich für ein populisti-sches Ausspielen

<indentation>(Unruhe)

von Leistung und Sich-wohl-Fühlen. Es ist eine Diskrimi-nierung von Grundschullehrerinnen in unserem Land. Die-se nehmen nämlich den Erziehungs- und Bildungsauftrag des Lehrplans der Grundschule ernst, der fordert, Grund-schulen als Lern- und Lebensorte auszugestalten, als Orte, an denen sich Kinder auch wohl fühlen können.

<indentation>(Beifall bei den Grünen und des Abg. Zeller SPD)

Dabei gibt es durchaus Reformprojekte, die wir unterstüt-zen, Frau Ministerin: zum Beispiel die frühere Einschulung gekoppelt mit dem „Schulanfang auf neuen Wegen“, die jahrgangsübergreifenden Klassen, bei denen das Prinzip gilt: differenzieren und fördern.

Wenn man sich allerdings anschaut, wie die Entwicklung verläuft, muss man sagen: Nachdem Sie sich bundesweit

damit profiliert haben, haben Sie das Interesse verloren. Das Reformprojekt dümpelt vor sich hin. Stattdessen hätte es ein Schlüsselprojekt für die Weiterentwicklung der Grundschule werden können. Nur 4 % der 2 500 Grund-schulen haben dieses Projekt tatsächlich durchgeführt.

Ich vermisse somit die Bereitschaft, Frau Kultusministerin, innezuhalten und auch einmal kritisch zu fragen: Mache ich, machen wir in diesem Bundesland eigentlich alles richtig? Man darf nicht immer nur sagen: Wir sind Spitze, wir können alles.

<indentation>(Beifall der Abg. Heike Dederer GRÜNE – Abg.
<indentation>Pfister FDP/DVP: Außer Hochdeutsch!)

Nachdenklichkeit, meine Damen und Herren, ist allerdings auch in diesem Hause, bei uns selbst, angesagt. Ich würde es für ein gutes Zeichen halten, wenn sich der Landtag ent-schließen könnte, eine Enquetekommission zum Thema „Weiterentwicklung von Schule und Unterricht“ einzuset-zen. Ich möchte daran erinnern – Frau Kollegin Rudolf hat dies ja bereits angesprochen –: Wir sind sehr zufrieden da-mit, dass wir in der letzten Legislaturperiode die Enquete-kommission „Jugend – Arbeit – Zukunft“ hatten. Sie hat genau zu der Erkenntnis geführt, dass ungefähr 20 % der Jugendlichen aus sozial benachteiligten Familien nicht mehr die Leistungen erbringen, die für eine berufliche Inte-gration notwendig sind. Die Jugendenquetekommission hat im Ergebnis ein Bündel von Maßnahmen empfohlen.
Unter anderem hat sie die Re-gierungskoalition endlich davon überzeugt, wie dringend
notwendig die Schulsozialarbeit in diesem Land ist. Sie hat dafür gesorgt, dass in Baden-Württemberg Jugendagentu-ren eingerichtet wurden, die den Jugendlichen helfen, den schwierigen Übergang von der Schule in den Beruf zu meistern. Deshalb, sage ich, würde es uns gut anstehen, zunächst ein-
mal genau hinzuschauen und zu klären: Wo liegen denn die Schwächen? Vor allem sollten wir aber klären: Welche ganz konkreten Handlungsperspektiven müssen aufgebaut werden, damit alle Jugendlichen, von den sozial benachtei-ligten bis zu den höchstbegabten, die Bildung bekommen und die Kompetenzen entwickeln können, die sie von ihren Potenzialen her mitbringen? Ich bedanke mich.

<indentation>(Beifall bei den Grünen und des Abg. Kaufmann
<indentation>SPD)

<poi_begin>Präsident Straub:<poi_end> Das Wort erteile ich Frau Ministerin Schavan.

<poi_begin>Ministerin für Kultus, Jugend und Sport Dr. Annette Schavan:<poi_end> Herr Präsident, meine sehr verehrten Damen und Herren! 1997 hat sich die Kultusministerkonferenz in Kon-stanz entschieden, künftig deutsche Schulen an internatio-nalen Vergleichsstudien zu beteiligen. In den letzten zehn Tagen habe ich mich an diese Situation, an die damalige Sitzung und die Wochen und Monate danach erinnert, und ich habe mich übrigens auch an manche schul- und bil-dungspolitische Debatte der letzten Jahre in diesem Haus erinnert.

<indentation>(Abg. Röhm CDU: Jetzt kommts!)

FIGURE B2: PLAIN TEXT REPRESENTATION OF PDF DOCUMENT

Appendix C – Topic Classification

Topic classification is crucial for my analysis. I have used both supervised and unsupervised machine learning methods to achieve this task. The rationale for combining these methods is to obtain a reliable classification of the topics in the corpus at a relatively low cost. Supervised machine learning methods allow the researcher to have more control over the classification task but require manual labelling of a subset of the data, which is labor- and time-intensive. I have adopted this method to classify the most important topic for my analysis: education. Conversely, unsupervised machine learning methods for topic classification have the advantage that they do not require any manual labelling and do not require the researcher to know all the topics of the corpus in advance, but are harder to interpret. I have used this method to classify the topics of all the speeches that I have classified as not being about education. In the following sections, I provide a brief description of the classification task separately for the supervised and unsupervised machine learning methods.

C1. – Classification of Education Topic

Supervised machine learning (SML) methods require labelled data to learn the relationship between the outcome of interest and the available explanatory variables. I have therefore instructed two research assistants to manually classify 48 plenary sessions for a total of 3,346 speeches. The sessions were picked randomly from each state to ensure representativeness of the labelled dataset. The selection of sessions was slightly adjusted to favor sessions that discussed education topics. Specifically, randomly selected sessions were discarded if the word "school" did not appear in the entire session. It is important to remind that plenary sessions tend to be quite lengthy and deal with plenty of issues. Thus, favoring the sessions in which education topics are discussed does not prevent other topics from being adequately represented. The research assistants classified each speech in a binary way: whether it is about education or not. For the purpose of my analysis, I instructed the research assistants to consider speeches as being about education if they concern any education-related topic at the elementary, primary and secondary school level. Higher education was not considered part of the education theme, as it has a different legal basis and tends to be mandated to different ministries. The research assistants classified 571 speeches about education, 17.1% of the total number of speeches classified, while the other 2,775 were classified as not being about education.

At this stage, I face a binary classification task. The aim is to learn the conditional expectation function $Y(X)$, where $Y \in \{0,1\}$ denotes a binary indicator of whether the speech is about education and X denotes a vector representation of the speech, that governs the relationship between the label and content of the speech. I will then use the estimation of such function to predict the label of the entire corpus. I transform the speeches into a vector representation in three steps. I first perform standard preprocessing steps such as removal of stopwords, punctuation and numbers. Then, I apply a term frequency-inverse document frequency ($tf-idf$) transformation of the entire corpus of speeches. In this case, I apply a standard $tf-idf$ transformation of each speech according to the following formula:

$$tf-idf_d \equiv \frac{c_{dw}}{\sum_{k \in d} c_{dk}} \times \ln\left(\frac{D}{\sum_{n \in D} \mathbb{I}(c_{nw}>0)}\right), \tag{C1}$$

Differently from the $tf-idf$ transformation used described in Section 3.5, in this case I do not perform a topic version of the $tf-idf$. The $tf-idf$ transformation of the documents upweights words that are specific to certain documents and downweights words that occur in many documents. I exclude words that occur in more than 50% of the documents and in less than 1% of the documents, as these terms are either too common or too rare. Given the limited size of the labelled dataset, I perform a further step to reduce the dimensionality of the explanatory variables. I implement the topic modelling algorithm Latent Dirichlet Analysis (LDA). LDA is a machine learning algorithm that identifies topics in corpora of texts in an unsupervised way based on the frequency with which words co-occur together. The crucial input parameter for this algorithm is the number of topics, which is unknown to the researcher and affects the broadness and interpretability of the topics. In this specific application, where I use LDA as a step for dimensionality-reduction purposes, I chose 15 topics as this was the number of topics that provided the highest accuracy in the classification task.

I then split the manually labelled sample into a train (80%) and test (20%) sample, stratifying by the binary outcome to ensure that both the train and test sample contain an equal share or speeches about education. I use a logistic regression classifier with 5-fold cross-validation and tune the hyper-parameters of the classifier using grid search over the type of penalty and strength of the regularization.[36] The best estimator is a logistic regression with an L1 type of penalty and a regularization hyper-parameter equal to 100.

[36] I have also tried other classifiers, such as Random Forest, Lasso, XGBoost and Gradient Boosting and achieved equivalent results in the classification task.

I report the evaluation metrics of the classification task in Table C1. Overall, figures show that the logistic regression achieves very good results in the classification of education vs non-education speeches. The F1 score, a metric which combines the precision and recall of the classifier, is close to 1, the maximum value for such metric.

I then use the machine learning model trained on the labeled dataset to make out-of-sample predictions for the entire corpus. I classify 18,701 speeches as being about education, or 8.9% of the speeches. I provide further descriptive evidence to corroborate the reliability of the classification task. In Table C2, I report the average number of times a set of words typical of the education context are mentioned in a speech. I select the terms “school”, “teacher”, “education”, and “lesson”. On average, the term “school” in column 1 is mentioned 8.29 times in speeches classified as being about education, while only 0.55 in non-education speeches. Similarly, the terms “teacher”, “education” and “lesson” are mentioned much more often in education speeches.

C2. – Classification of Other Topics

In cases where at least some of the topics in the corpus are unknown to the researcher and the corpus has not been manually labelled, topic models offer a fast and cheap solution to the classification task. Topic models are latent variable models that exploit the correlations among the words and latent semantic themes. For the classification task in this paper I apply the correlated topic model (Blei and Lafferty 2007)[37] to the corpus after the standard pre-processing steps described in the previous section. The correlated topic model (CTM) has the advantage over the more common Latent Dirichlet allocation (LDA) topic model of explicitly modeling the correlation between the latent topics in the corpus. I take advantage of this feature to aggregate similar topics, as topics do not be overly narrow to compute the polarization measure.

Like all topic models, the key tuning parameter of the CTM is the number of latent topics *K.* The outcome of the CTM depends largely on this parameter, which is mainly set depending on the size of the corpus, prior knowledge of the researcher about the corpus, and the downstream task the researcher wants to achieve. A low *K* might induce the CTM to aggregate unrelated topics, whereas a large *K* might split a single topic into excessively narrow sub-topics. In my setting, the corpus is relatively large, but I am not interested in narrowly defined topics. Ideally, topics should be of similar size to the

[37] I used the R package *stm,* which enables a fast implementation of the correlated topic model (for further details, see Roberts, Stewart, and Tingley (2019))

education topic classified using the SML algorithm in the previous step, which is around 9% of the corpus.

Once the CTM with K topics is estimated, the researcher needs to assess the outcome of the model by manual inspection of the identified topics and the provided metrics. A CTM with 30 topics provided good outcomes in terms of interpretability of the topics. I report the estimated topics in Table C3, with the 5 most relevant words for each topic and the manually assigned label. I aggregate topics which are either semantically similar or display a high correlation and identify 11 distinct topics. This step ensures that the size of the topics is similar to the education topic. I provide a graphical representation of the correlation among topics in Figure C1. Further, the heatmap places correlated topics next to each other and clusters of topics can be identified by looking at the dendrogram built on top of the heatmap. For example, the heatmap places topic 16 and 17 next to each other, as they display a high level of correlation. These topics concern discussion about housing and infrastructures, as it can be inferred from the most representative words reported in Table C3. Thus, I aggregate these topics into a single topic labelled "Housing, Infrastructure and Transportation". One topic, labelled "Other", does not have a clear interpretation and is not highly correlated with any of the other topics.

In principle, I could avoid aggregating topics while still obtaining large enough topics by using lower K, as the average size of the topics strictly decreases in K. However, using lower K led to worse performance of the CTM in terms of interpretability of the topics. Further, the final size of the topics was largely heterogenous, with some very small-sized topics and some relatively large topics. Nonetheless, it is reassuring that the number of topics K set at this stage does not affect my analysis in a substantial way, as I show in Table A6 in Section 5.5.

For each speech in the corpus, the CTM provides the estimated weight of each latent topic, with weights summing up to one. I assign the topic with the largest weight to each speech, which allows me to obtain a categorical classification of the corpus.

As a final step, I combine the estimation of the education topic described in the previous section and the topics obtained with the CTM. I assign the topic "education" to all the speeches that the SML algorithm classifies as being about education. For the remaining speeches, I assign the aggregated topics reported in Table C3. As expected, the CTM also identifies two topics that are clearly about education, namely topic 9 and topic 18, since I applied the CTM to the entire corpus.[38] However, I do not assign the education topic identified by the CTM to any speech, as this *de facto* contradicts the more reliable SML classification, which predicted such speeches as not being about education. Given the high correlation between the SML classification of speeches about education and the CTM education topics (see footnote 18), such conflicts are rare: they concern less than 1% of the speeches. In these cases, I assign the second largest topic to these speeches instead. I report the share of aggregated topics for the main analysis in Figure A1. It can be noted that the size of the education topic lies at the median of the distribution.

TABLE C1—CONFUSION MATRIX – CLASSIFICATION OF EDUCATION SPEECHES

	Precision	Recall	F1 Score	Support
	(1)	(2)	(3)	(4)
Non-Education	0.96	0.98	0.97	556
Education	0.88	0.81	0.84	114
Weighted Average / Total	0.95	0.95	0.95	670

Note: Confusion matrix of the classification of non-education and education speeches using a Logistic Regression. In column 1, I report the precision rate of the classification task, in column 2 the recall rate, in column 3 the F1 score. In column 4, I report the sample size, of both categories and the total sample size of the test sample.

TABLE C2—AVERAGE FREQUENCY OF EDUCATION TERMS

	School	Teacher	Education	Lesson
	(1)	(2)	(3)	(4)
Non-Education	0.55	0.08	0.72	0.06
Education	8.29	2.82	3.95	1.77

Note: Average frequency of the terms "school" (column 1), "teacher" (column 2), "education" (column 3), and "lesson" (column 4) by speeches classified as being *not* about education and speeches about education in the entire corpus (210,006 speeches). The original German terms searched are "Schule", "Lehrer", "Bildung", and "Unterricht".

[38] Alternatively, I could apply the CTM only to those speeches classified as not being about education by the SML algorithm. However, doing so did not resolve the issue in this case, as often the CTM identified an education-related topic nonetheless. Note that this might not only be due to the SML classification being imperfect, but also to the fact that speeches often touch upon different topics. For the purpose of my analysis, the SML only classifies speeches in a binary way, whereas in reality the distinction is fuzzier than such classification might suggest. Thus, speeches classified as not being about education might still reference to education-related issues. A further reason to apply the CTM to the entire corpus is that it gives me the opportunity to compare the SML classification of the education topic with the CTM classification of education topics. Reassuringly, the SML classification of the education topic is highly correlated with the education topics identified by the CTM (.71, p-value <.01). This increases the credibility of my classification task.

TABLE C3—TOPIC LABELS AND TOP 5 FREX WORDS

Topic Number	Example FREX Words	Topic
1	museums, culture foundation, cultural policy, orchestra, theater	Higher Education, Culture
2	resolution recommendation, bill, bill, bill, legal	Lawmaking Process
3	supplementary budget, debt, net neuverschuldung, net borrowing, debts	Economy, Labor Market, Public Finance
4	national park, waters, water framework directive, national parks, forest owner	Agriculture, Environment, Energy
5	koch, committee of inquiry, roland, prime minister, investigative committee	Government Operations
6	preventive detention, rehabilitation, juvenile detention, prisoners, jva	Security, Crime, Defense
7	(university) students, courses, study places, higher education law	Higher Education, Culture
8	state government, called, weiss, measures, big	Government Operations
9	religious education, Pisa study, pisa, integration policy, comportment grades	Education
10	energies, renewable, renewable, nuclear energy, power plants	Agriculture, Environment, Energy
11	application, to discuss, application, think, points	Lawmaking Process
12	minimum wage, minimum wages, wages, collective agreements, moonlighting	Economy, Labor Market, Public Finance
13	gender, mainstreaming, men, women politics, women	Social Welfare, Healthcare, Equality
14	people's union, war, soldiers, sed, ddr	Security, Crime, Defense
15	municipal, municipal, circle-free, municipal financial reform, connection principle	Local Politics
16	renter, urban redevelopment, housing company, housing industry, housing market	Housing, Infrastructure, Transportation
17	temple courtyard, bust, schönefeld, bbi, lb	Housing, Infrastructure, Transportation
18	school board, school boards, class failure, principal, student numbers	Education
19	civil servants, officials, district courts, Officer, fire brigades	Security, Crime, Defense
20	agricultural policy, genetically, bse, modulation, animal meal	Agriculture, Environment, Energy
21	care insurance, insured, dependent, patients, statutory health insurance	Social Welfare, Healthcare, Equality
22	savings banks, broadcasting, broadcasting amendment state treaty, ard, state banks	Economy, Labor Market, Public Finance
23	federal traffic route plan, road charge, freight transport, passengers, long-distance	Housing, Infrastructure, Transportation
24	tell, rode, talked, outside, people	Government Operations
25	apprenticeship places, apprenticeship fee, labor market policy, long-term unemployment	Economy, Labor Market, Public Finance
26	answer, answer, exam, answered, documents	Other
27	child poverty, child protection, day care centers, family policy, childminders	Social Welfare, Healthcare, Equality
28	hardship commission, petition committee, petitioners, petition commission, non smoking protection	Security, Crime, Defense
29	economic promotion, dockyard, structural change, north hesse, biotechnology	Economy, Labor Market, Public Finance
30	right-wing extremism, extremism, far right, right-wing extremist, right-wing extremists	Security, Crime, Defense

Note: List of 30 topics identified by the CTM model. For each topic, I report the five most representative words according to the Frequency-Exclusivity (FREX, see Roberts, Stewart, and Tingley (2019)) metric and the manually assigned topic. The FREX words have been translated into English using the *deep translator* package in Phython.

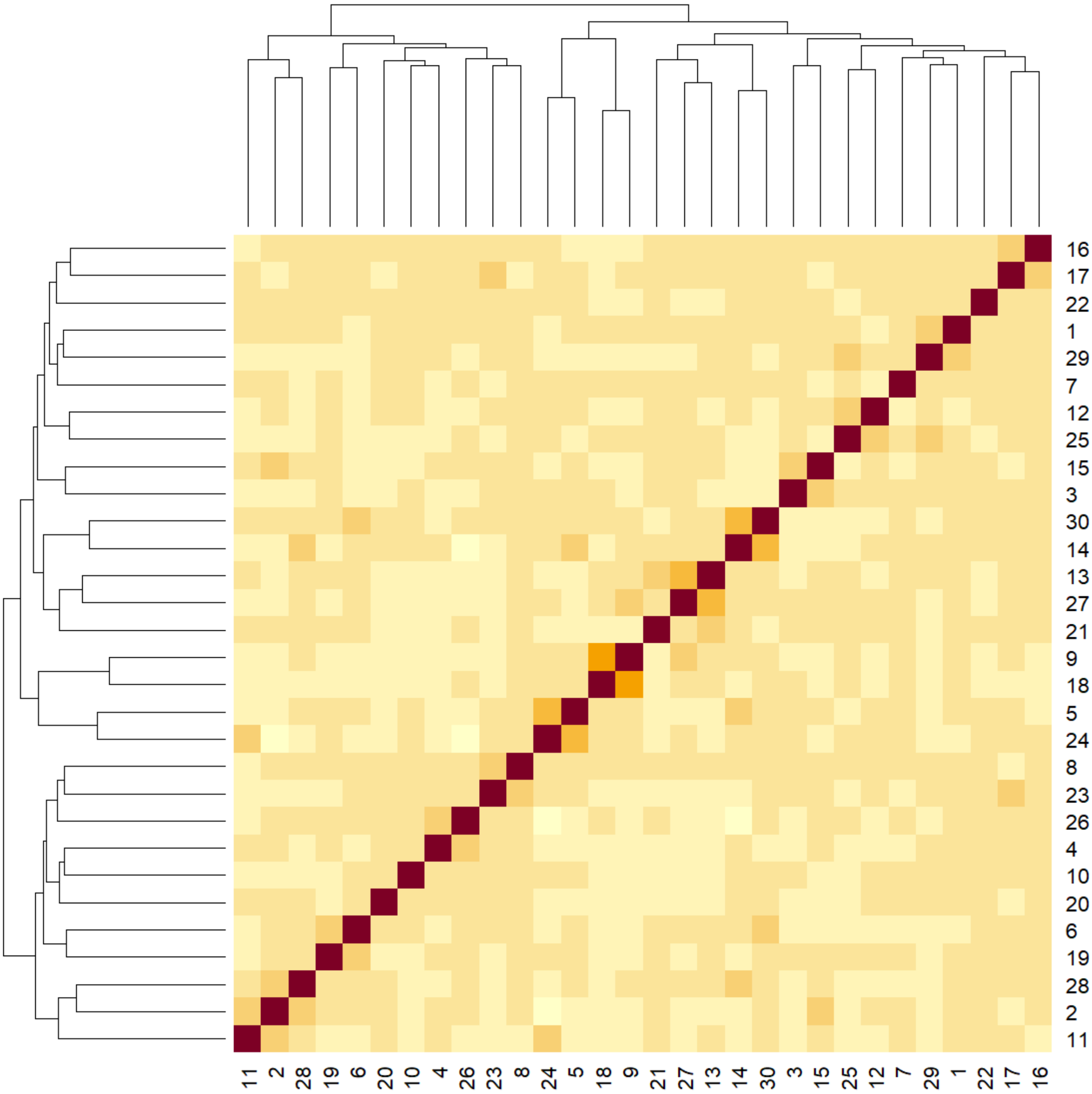


FIGURE C1: ORDERED HEATMAP OF TOPIC CORRELATION

Note: The ordered heatmap depicts correlation between each topic and all the other topics. Topics are reordered using a clustering algorithm which arranges topics by similarity, thus placing more correlated topics next to each other. The overlayed dendrogram arranges cluster of topics by their correlation with each other. Darker colors indicate higher correlation.