# Topics as Outcomes: Using Structural Topic Models to Measure Policy Diffusion

NaLette Brodnax
Georgetown University

Sarah James
Harvard University

Prepared for the 35th Annual Meeting of the Society for Political Methodology, July 18-21, 2018, Provo, UT

**Abstract**

For most analyses of policy diffusion, the outcome of interest is a dichotomous measure of adoption, derived by comparing the major provisions of legislation. This approach has limited utility for comparing a large number of cases where states adopt only part of another state's policy, or for assessing the magnitude of similarity between states. To address these limitations, we propose a measure of adoption based on the prevalence of topics across many pieces of legislation. Specifically, we use structural topic modeling to estimate the prevalence of particular policy topics, as they appear in legislation across states and over time. We apply this method to the range of education policies that US state legislatures introduced following the Race to the Top education grant competition. From 2010 to 2012, the US Department of Education awarded $4 billion in competitive grants for state-level reforms meeting numerous criteria for innovation. To analyze the extent to which these innovations appeared in state policies, we constructed a corpus from a query of over 40,000 education bills introduced by state legislatures from 2009 to 2016. Our methodology proceeds in two stages. First, we estimate the prevalence of education policy topics for the set of documents, and compute the mean prevalence of each topic for each state. Second, we construct a multilevel model to estimate policy diffusion, using the state-dyadic difference in topic prevalence as the outcome. We include demographic, political, and geographic similarity covariates for each dyad, and a set of characteristics for each state. We find that evaluation policies spread from states who were awarded competitive grants to those that did not, suggesting that states learn from rather than compete with states for the grant funding.

# 1 Introduction

Policy diffusion is the process through which policies adopted by a governmental jurisdiction are influenced by other jurisdictions within the system (Berry and Berry, 2014). Studies of policy diffusion in the United States address a wide range of questions pertaining to the extent to which certain policies spread over time, states' motivation for adopting the policies of other states, and the aspects of the federalist system that facilitate diffusion. States might be influenced by a range of mechanisms, such as learning, imitation, normative pressure, competition, or coercion, and draw on the experiences of states sharing their geography, ideology, or other characteristics.

Policy diffusion scholars face methodological challenges with both data collection and measurement. These include difficulty identifying the universe of observations (Nicholson-Crotty, 2009), gathering comparable data on policies over time (Shipan and Volden, 2006), and constructing valid measures of the phenomena. Policy diffusion, like other forms of diffusion, is characterized by the cumulative distribution of adoptions taking on an S-shaped curve, as shown in Figure 1. For most policy diffusion studies, the units of analysis are states, state-years, or state-dyad-years, and the outcome of interest is a binary measure of adoption. The adoption measure is derived by comparing the major provisions of legislation and assigning a value of one if a state adopts a particular policy (e.g., see Berry and Baybeck (2005), Shipan and Volden (2006), and Nicholson-Crotty (2009)), or if one state adopts a policy that is similar to another state (e.g., see Volden (2006), Shipan and Volden (2014), and Volden (2015)). Comparing policies to assess similarity is time consuming, requires subjective expertise, and is difficult to replicate (Volden, 2015). Further, relying on a binary measure of adoption discards information about the magnitude of similarity between two policies (Berry and Berry, 2014).

We address these challenges through automated text analysis, which can be used to process, categorize, and analyze political corpora (Grimmer and Stewart, 2013). Further, we propose the use of structural topic modeling (STM) to create a continuous outcome measure of

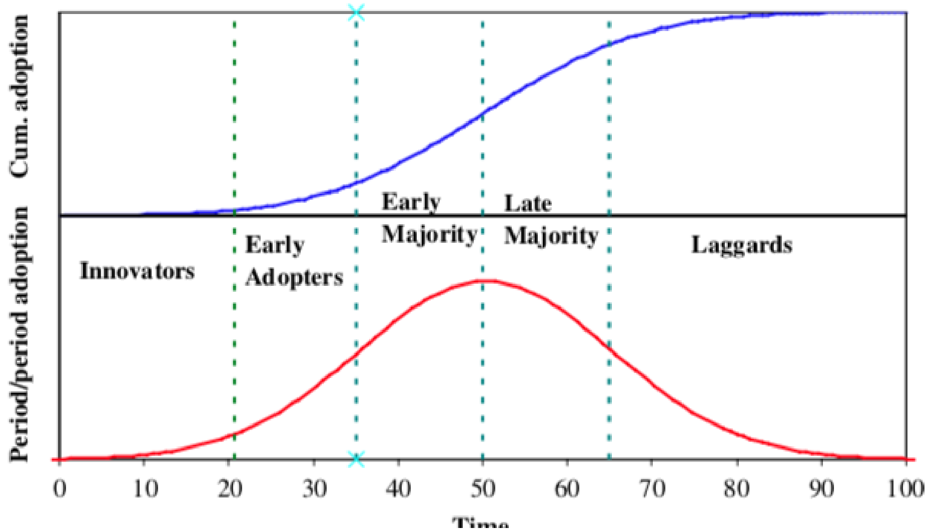


Figure 1: Stylized Diffusion Curves (Meade and Islam, 2006)

policy adoption based on the proportion of words in each document corresponding to a given topic. Structural topic models are a class of topic models that incorporate document-level metadata, such as author and date, as covariates. The STM generates the proportion of each document corresponding to a set of topics, called topic prevalence. We use topic prevalence as the measure of adoption, and compute the difference in topic prevalence between each state-dyad as the basis for estimating diffusion. In a similar application (Gilardi et al., 2017) examine the issue definition stage of the policy process and uses STM to analyze text from newspaper articles.

We apply this method to the range of education policies that US state legislatures introduced following the Race to the Top education grant competition (US Department of Education, 2013). From 2010 to 2012, the US Department of Education awarded $4 billion in competitive grants for state-level reforms meeting numerous criteria for innovation. To analyze the extent to which these innovations appeared in state policies, we queried a corpus of over 40,000 education bills introduced by state legislatures from 2009 to 2016. Our methodology proceeds in two stages. First, we estimate the prevalence of education policy topics for the set of policies, and compute the mean prevalence of each topic for each state. Second, we construct a multilevel model to estimate the diffusion of teacher evaluation policies, using the state-dyadic difference in topic prevalence as the outcome. We include demographic, political, and geographic similarity covariates for each dyad, and a set of characteristics for each state.

## 2 Teacher Evaluation Policy Reform

Public education has traditionally been the responsibility of state governments, which are better able to adapt systems to local conditions. However, the federal government has dramatically increased its role in education policy over the last two decades (Cohen-Vogel, 2005; Lavigne and Good, 2014). Table 1 provides a summary of federal education policy interventions. While the policies target different actors (states, districts, schools, etc.), they share a common goal of increasing teacher *effectiveness*, use a variety of incentives to achieve this goal, and fall into two categories: social regulation and grants. Generally, social regulation comprises laws and rules that seek to control behaviors that impact public health, safety, welfare, or well-being (Salamon and Elliott, 2002).

The No Child Left Behind Act of 2001 (NCLB), a wide-reaching policy intended to reform public education by 2014, had numerous consequences for states. One of the laws provisions required states to provide highly qualified teachers for all students and ensure an equitable distribution of teachers for all populations of students (American Institutes for Research, 2014). Another provision spurred a coordinated effort to collect achievement data for all of the nations public schools. The process of collection such a massive amount of data gave policymakers a better understanding of the problem and set the stage for future policy interventions. Though NCLB quantified the failure of American schools, its primary mechanism of reformmandatefailed to raise achievement as intended (Lavigne and Good, 2014).

Salamon and colleagues argue that the effectiveness of social regulation depends on (1) the reasonableness of what is required, the capacity and willingness of regulated entities to comply, and (3) the adequacy of resources for enforcement (Salamon and Elliott, 2002). As

Table 1: Summary of Federal Policies Affecting Teacher Evaluation

| Year | Type | Policy | Provisions Related to Teachers |
|---|---|---|---|
| 2001 | Regulation | No Child Left Behind/Elementary and Secondary Education Act Reauthorization | Defined a "highly qualified" teacher; required equitable distribution of highly qualified teachers |
| 2004 | Regulation | Individuals with Disabilities Education Act Reauthorization | Required that special education teachers meet NCLB guidelines |
| 2006 | Grant | Teacher Incentive Fund | Many early TIF grants used to put teacher evaluation systems in place |
| 2009 | Grant | Investing in Innovation | Aims to "promote practices, strategies, or programs to increase the number and percentage of effective teachers and school leaders" |
| 2009 | Grant | School Improvement Grants | Schools encouraged to "develop comprehensive teacher and leader effectiveness systems" |
| 2010 | Grant | Race to the Top (via the 2009 American Recovery and Reinvestment Act) | Encouraged applicants to collect student data and tie it to teacher performance ratings |
| 2010 | Regulation | NCLB/ESEA Flexibility Waivers | In order to receive a waiver, states had to have plans for "teacher and leader evaluation systems that include growth in student achievement" |
| 2012 | Grant | District Race to the Top | Encourages "longitudinal data systems that can measure and report on student growth, student achievement, and instructional needs" |

Source: American Institutes for Research (2014)

stipulated in the law, states were required to show adequate yearly progress or face severe sanctions. However, Congress provided states with little guidance on how to make adequate yearly progress and little funding to support reform initiatives. Few states were able to comply with the law and the federal government had limited capacity to sanction states with school closures.

In response to the perceived failures of NCLB, more recent federal policies offered competitive grants to states in order to push reform. These grant competitions, several of which aim to influence teacher effectiveness, differ from social regulation in important ways. First, grants are less coercive than mandates, but still succeed at stimulating the desired behavior (Salamon and Elliott, 2002). Second, enforcement is built in: states that fail to enact reform also fail to win competitive grants. Finally, grants do not require states to fund initiatives without adequate resources, although some grants require states to match the funds from their own coffers.

## 2.1 Race to the Top

Race to the Top (RTTT), a competitive grant program created by the US Department of Education in early 2009, is one of the largest and most visible grant competitions targeting education reform. As part of the American Recovery and Reinvestment Act, the Department received $4.35 billion to support state-level education reforms, particularly those innovative strategies that are most likely to lead to improved results for students, long-term gains in school and school system capacity, and increased productivity and effectiveness (US Department of Education, 2013; US Congress, 2009). States were invited to apply for the funding and then evaluated according to a number of selection criteria. The application guidelines essentially require states to adopt policies *prior to* submitting an application, including policies that mandate the collection and maintenance of student performance data.

While Race to the Top did not require a specific use for the student data collected, it encouraged states to tie teacher evaluation to student achievement data. States whose grant applications were most competitive not only mandated data collection, but also required schools to use the data to improve teacher effectiveness. Forty states and the District of Columbia applied for RTTT grants, with 19 receiving awards.

States moved quickly to pass laws targeting teacher evaluation, shifting from the NCLB emphasis on teacher quality to an emphasis on teacher effectiveness. However, evaluation policy requirements vary widely with respect to the type of evaluation system and the use of student achievement data within those systems. Twenty-seven states provided guidelines and optional evaluation models, 11 states and the District of Columbia mandated the use of a single, state-wide evaluation system, ten states provided a state-wide system but allow districts to opt out, and two states provided no state-wide specifications (Doherty and Jacobs, 2013). Despite considerable variation with regard to evaluation systems, state legislatures made clear that any system should tie teacher performance to student outcomes: forty states plus the District of Columbia have adopted legislation to that effect. Of those, 17 require schools to weight student achievement data by a specific percentage, ranging from five percent to at least 50%. The remaining states legislation requires that student achievement data be includedsome specifying that it be included as a "significant factor.

Numerous studies show that teacher quality is the most important school-level predictor of student achievement (Rivkin et al., 2005; Heck, 2007). For example, one study estimates that improving the bottom 5-8% of teachers would move the United States to the top of international rankings (Hanushek, 2011). However, given limited attention and limited resources, the political choice to focus on teacher evaluation comes at the cost of other reform opportunities. Further, some argue that an individual teachers ability to influence student growth pales in comparison to socioeconomic factors. Student achievement is most highly correlated with factors outside the school, such as poverty and racial segregation (Rothstein, 2004; Orfield and Lee, 2005). Changes in teacher evaluation policy also affect teacher compensation, promotion, tenure, discipline, and dismissal. In this study, we seek to understand the role of diffusion in the spread of teacher evaluation policies.

# 3 Policy Diffusion: Competition or Learning?

The authors of the Race to the Top guidelines explicitly included policy diffusion as part of the overall vision for reform. Proponents of the program hoped that rewarding innovation would create models for effective education reform that other states would then adopt in hopes of earning future RTTT awards. The executive summary of the RTTT application packet states this directly, saying that successful states "will offer models for others to follow and will spread the best reform ideas across their States, and across the country (US Department of Education, 2013).

The RTTT grant application delineated six categories of criteria against which states would be evaluated: state success factors, standards and assessments, data systems to support instruction, great teachers and leaders, turning around lowest achieving schools, and general selection criteria. This list also contained each categorys respective point allotments in the decision-making process with the "Great teacher leaders" accounting for almost a third of the total points, more than any other category. In the spirit of encouraging innovative practices to improve student outcomes, rather than mandating specific reforms, RTTT outlined what it considered to be essential qualities of a teacher evaluation system. Successful states in the application would have a teacher effectiveness tool in which student growth played a "significant role, but also included multiple additional measures of teacher effectiveness.

The multi-staged design of the RTTT application process made it possible for states to learn from one another. RTTT held three application rounds —due in January 2010, June 2010, and November 2011 —and states could participate in each round until they received a grant. Once a state successfully earned a grant, it was no longer eligible to apply for additional funds. To encourage emulation of grant-recipients, RTTT also mandated that all application materials be posted online shortly after the completion of an application round to ensure that other states could examine the programmatic components of successful state applications. Furthermore, states were also required to show evidence that the legislature and executive branches had approved any relevant policy changes before the submission of the grant application. This requirement suggests that attempts at creating new teacher effectiveness measures should appear in legislative documents if the state hoped to win any RTTT grant funding.

The design of and response to the announcement of RTTT leads us to believe that the program created conditions ripe for policy diffusion. First, all but four states (Alaska, North

Dakota, Texas and Vermont) submitted applications during at least one of the three rounds. Second, the application window was narrow —from June 2010 to November 2011 —and this quick turn-around for developing a reform plan likely encouraged states to look to their peers for programs worth emulating, rather than developing reform initiatives from scratch. Third, the incentives were substantial; grants ranged from 17 million to 700 million dollars, with the average grant award hovering just below 220 million dollars. Fourth, the RTTT application contained a rubric with explicit criteria against which sates would be judged. In other words, there was little doubt about the types of reforms that the program rewarded, though the exact strategies for implementing the reforms were left up to the states. Finally, as mentioned above, the leaders of the RTTT process announced their intention to encourage policy diffusion, which suggests that states looking to win funds in the second and third rounds would have looked to successful recipients in prior rounds for inspiration on how to enhance their application.

We consider two possible diffusion mechanisms: competition and learning. If states adopted teacher evaluation reforms in order to compete for grant funds, then we expect to see a spike in reforms following the announcement of the program. If states sought to emulate states who were awarded grants, then we expect to see a significant relationship between the sender state receiving a grant in the prior period, and the receiver states' policies moving closer to that of the sender state. Though RTTT competitions officially ended in 2011, our study examines all policies related to teacher evaluation through 2015. The political pressure to address educational inequality and lackluster outcomes did not end along with the conclusion of RTTT; states therefore plausibly still faced incentives to emulate successful policies.

## 3.1 Topics as Outcomes

Legislatures divide their policymaking efforts across a set of topics due to limited attention. Our empirical model assumes that each bill that they produce is drawn as a mixture of a fixed number of topics. We estimate this mixture for each bill using structural topic modeling (STM), a text analytical method that categorizes each document into a mixture of topics based on features of the text as well as document-level metadata (Roberts et al., 2016). Topic modeling requires few assumptions and can be automated to analyze a large volume of documents at a low cost and in relatively little time (Quinn et al., 2010; Grimmer and Stewart, 2013). Structural topic modeling has the additional benefit of allowing the analyst to weight the mixture of topics on observed document covariates. These covariates can be incorporated into the model's document-topic proportions, called *topic prevalence*, or topic-word rates, called *topical content*. For this component of the analysis, our quantity of interest is topic prevalence.

Given a set of documents $D$ indexed by $d$, words at position $n$ within each document, and a vocabulary of distinct words $V$, the STM generates each document from a mixture of topics $K$, indexed by $k$. The distribution of topics is informed by the set of document covariates $X$. Document $d$ is generated by sampling topics and words as follows:

1. Draw topic $z_{d,n} \sim \text{Multinomial}(\theta_d)$

2. Draw word $w_{d,n} \sim \text{Multinomial}(\beta_{z_{d,n}})$

where $\theta_d$ is the $D$x$K$ matrix of document-topic proportions, and $\beta_{z_{d,n}}$ is the $K$x$V$ matrix of word-topic frequencies.

Structural topic modeling treats each document as a bag of words, ignoring the order of the words. Given this loss of information, the analyst must take care in validating and interpreting the model. The $\theta$ matrix indicates the probability that a given document was generated from topic $k$. Parameters from this $D$x$K$ matrix can be used alongside the document metadata to aggregate topic prevalence across units. We calculate the mean topic prevalence for each state $i$ and year $t$ as the outcome of interest. We interpret mean topic prevalence for a given state as a summary of the extent to which a state adopted language associated with particular policy innovations. The $\beta$ matrix comprises the content of the model, which can be used to learn what each topic contains and how topics are related. We use parameters from this $K$x$V$ matrix to assess the validity of topics.

The workflow for analyzing text data involves four steps: (1) acquire, (2) prepare, (3) estimate, and (4) validate (Grimmer and Stewart, 2013).

### 3.1.1 Acquisition

Legiscan LLC publishes a public web application programming interface (API) providing access to legislation in all 50 states since 2009, including bill status, sponsors, full text, roll call records, and other information[1]. Using the API, we identified 49,697 bills containing the term "education" and 9,628 containing the term "teacher." In order to narrow the scope of the analysis and limit variation in the number of topics, we searched for documents containing the term "teacher" in proximity to "quality" or "effective." This search produced 2,289 results for unique bills. We acquired all versions of text for each bill, yielding 5,296 documents.

Legiscan provides documents in the original format as posted on the state legislature's website: either HTML, PDF, or Microsoft Word format. We converted the HTML and Microsoft Word documents to plain text using Pandoc. For PDF documents, we extracted the plain text using Python, Ghostscript, and Tesseract, an open-source optical character recognition engine. We used the **stm** package (Roberts et al., 2017) for preparation, estimation, and validation.

### 3.1.2 Preparation

Pre-processing involves combining the documents into one file and modifying the text to reduce the dimensionality of the data. Once all documents were converted to plain text and combined into a corpus, we converted all words to lower case, removed numbers, and removed punctuation. Second, we removed any text added by Legiscan, a set of common stopwords, and the following domain-specific stopwords: shall, may, section, and subsection.

[1]See https://legiscan.com/

We stemmed all words for ease in aggregating words with similar meaning.

Prior to analyzing the corpus, we removed all bills containing "budget" or "approp" in the title. By excluding budget legislation, we eliminate a large source of variation in the number of topics. This left a sample of 923 bills that were introduced, engrossed, or enrolled by 38 states from 2009 to 2016. The **stm** package generates a vocabulary of distinct terms, first eliminating terms that appear only once in a document as well as documents containing only sparse terms. This yielded a final corpus of 921 bills with 588,841 words and a vocabulary of 26,281 terms. As metadata, we included chamber (House or Senate), date, and state of origin.

### 3.1.3 Estimation

We fit a 40-topic prevalence model as a function of state and chamber dummy variables, a splines function for the date, and an interaction between state and date to account for legislatures that meet in only odd years. Table 2 provides a summary of ten topics we identified and labeled, with mean prevalence and the most frequent words. As expected, the Evaluation topic includes stems related to teachers, performance, and effectiveness. Mean prevalence across states is 3.5%. Other topics appeared more prevalent, such as Licensing (6.74%), Employment (5.99%), and Reading (4.42%).

Table 2: Selected Topics and Top Word Stems

| Topic | Prevalence | Top Word Stems |
|---|---|---|
| Evaluation | 3.53% | school, evalu, district, perform, teacher, effect |
| Leave | 2.89% | sick, militari, leav, unus, absenc, credit |
| Retirement | 1.39% | retire, attain, age, adjust, beneficiari, allow |
| Licensing | 6.74% | teach, profession, mentorship, licensur, knowledg, induct |
| Income | 0.65% | gross, taxpay, taxable, incom, unitari, depreci |
| Charter Schools | 1.49% | charter, school, public, author, provid, state |
| Funding | 2.82% | scholarship, wherea, forgiv, shortag, tuition, repay |
| Reading | 4.42% | profici, onlin, defici, english, intervent, read |
| Employment | 5.99% | tenur, probationari, hear, perman, absenc, decis |
| Childcare | 1.47% | daycar, child, home, care, offend, childcar |

#### 3.1.4 Validation

To validate the topics, we consider their semantic validity, "the extent to which each category has a coherent meaning and the extent to which the categories are related to one another in a meaningful way" (Quinn et al., 2010). Semantic validity is maximized when the most probable words in a given topic are exclusive and frequently co-occur together, allowing the analyst to distinguish one topic from another. Semantic validity is influenced by the number of topics, assumed to be constant across documents. To assess models and topics, we use a metric known as FREX (Bischof and Airoldi, 2012) that balances frequency and exclusivity. We examined models with varying numbers of topics, eventually choosing 40 based on FREX scores. We also examined the semantic coherence and exclusivity for selected topics, as shown in Figure 2. The Evaluation topic exhibits the highest exclusivity and semantic coherence, relative to other topics.

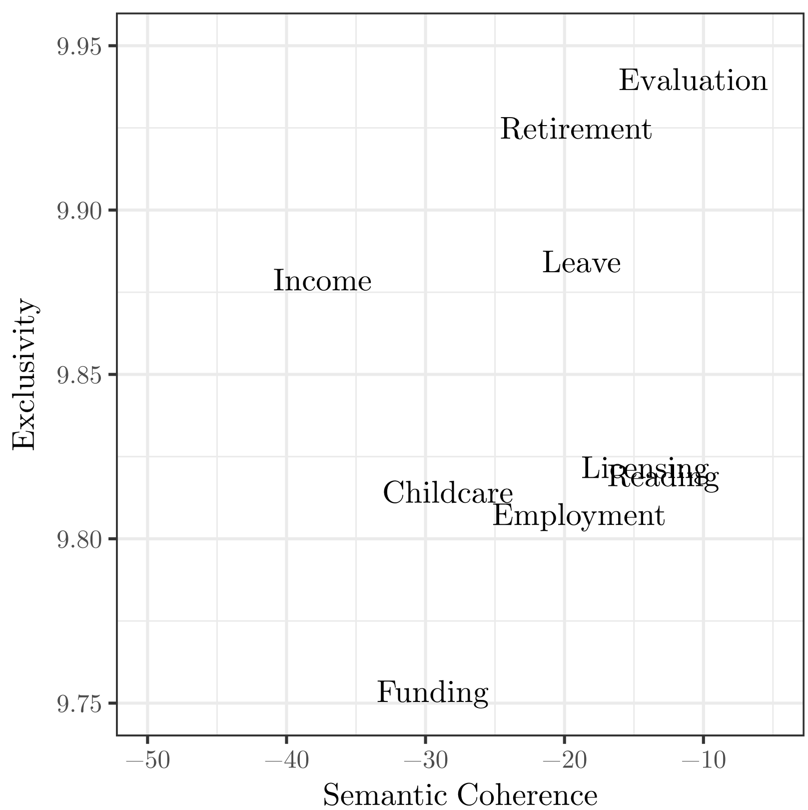


Figure 2: Topic Exclusivity and Semantic Coherence

For predictive validity, we can compare topic prevalence over time to see if it corresponds to known events (Quinn et al., 2010). Figure 3 shows the average monthly prevalence for the Evaluation topic over time. Given numerous federal programs aimed at teacher evaluation and effectiveness, Evaluation prevalence increases overall over the eight-year time span. In comparison to the period from 2009 to 2010, Evaluation spikes considerably during the application windows shown in red for each phase of Race to the Top.

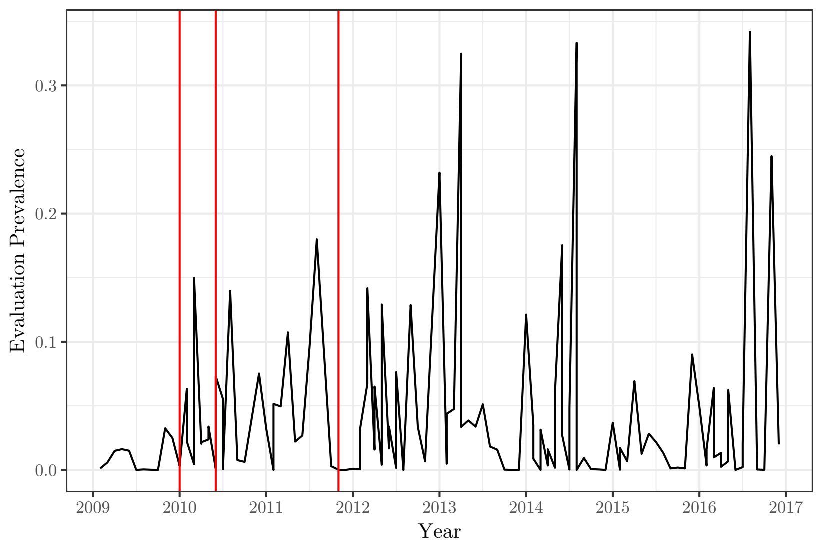


Figure 3: Evaluation Prevalence, 2009–2016

## 3.2 Policy Diffusion

In Figure 4, we show Evaluation prevalence by state for 2009 and 2011, the years immediately preceding and following the first wave of RTTT applications. In 2009 (hollow markers), just 10 states introduced legislation related to teacher evaluation. In contrast, 22 states introduced legislation in 2011. Some states' prevalence increased during the period, including California, Illinois, Michigan, and Ohio, while two states, Vermont and Washington, exhibited a decrease in prevalence. Most states' prevalence falls below five percent, but there is considerable variation across states. To assess relationships *between* states, we turn to the diffusion model.

To analyze the diffusion of a specific topic $k$ across states over time, we compute the

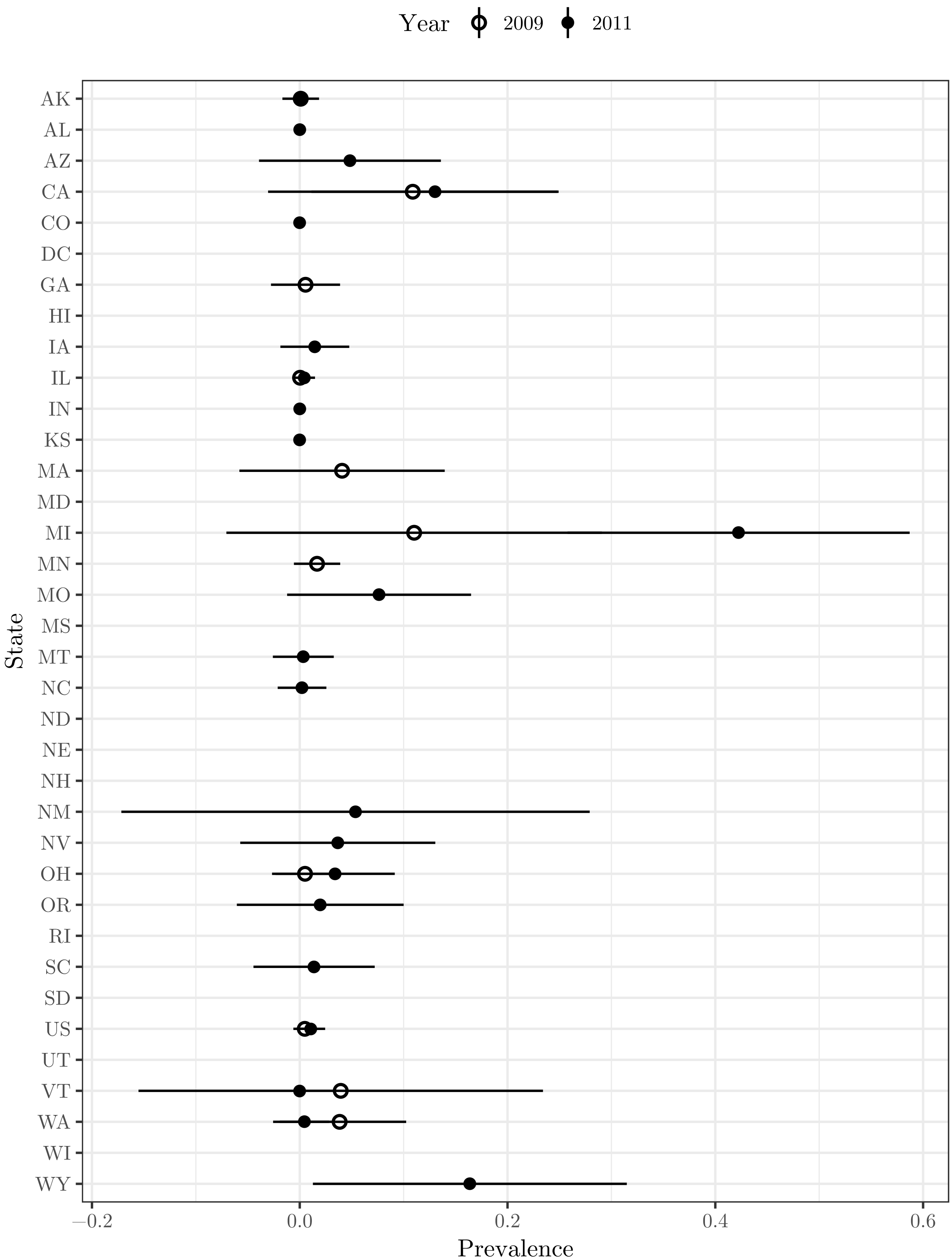


Figure 4: Teacher Evaluation Topic Prevalence by State, 2009–2011

mean prevalence for each state in each year. We use a directed dyadic approach, for which the unit of analysis is a pair of states (Volden, 2006; Gilardi and Fuglister, 2008). In the typical specification, the outcome variable is coded one when the receiver state adopts the same policy as the sender state. In contrast, we construct the outcome as the difference in topic prevalence between the receiver and sender state. To address a lack of independence between observations for each state-dyad over time, we specify a multilevel model, which allows estimates to vary across clusters and accounts for the complex structure of the data generation process (Gelman and Hill, 2007). We include a random intercept for each state and year:

Given receiver state $i$ and sender state $j$,

$$
\begin{aligned}
y_{ijt} &\sim N(\alpha_i + \alpha_j + \alpha_t + X\beta) \\
\alpha_i &\sim N(0, \sigma^2_{\alpha_i}) \\
\alpha_j &\sim N(0, \sigma^2_{\alpha_j}) \\
\alpha_t &\sim N(0, \sigma^2_{\alpha_t})
\end{aligned}
$$

where where $y_{ijt} = \mu_{it}^{\hat{\theta_k}} - \mu_{jt}^{\hat{\theta_k}}$ and $X$ is a vector of characteristics for $i$, $j$, and $ij$.

### 3.2.1 Data

As with all policy processes, policy diffusion does not occur in a vacuum. A state's history, political dynamics, past successes within the given policy area all affect the likelihood of adopting new policies piloted in other states. We include several covariates in our model to account for these factors.

First, we include several school-based measures. The availability of resources and the need for demand for educational services plausibly influences the pressure policymakers feel to make adjustments to existing practices, including teacher evaluation. Therefore, we include measures of both the student population size by state and the number of full-time employees designated as teachers.

We also include data on past student performance, as measured by math and reading proficiency on state standardized tests, and the ratio of highly qualified teachers present in each state. We rely on the No Child Left Behind definition of “highly-qualified,” which states that teachers must have a bachelors degree, full licensure from the state, and demonstrated competency in their core subject area, because it is a nationally recognized metric for which there is comparable state data. Furthermore, the highly-qualified designation offers an albeit-rough measure of teacher quality upon entrance into the classroom. The effectiveness of past policies influences the degree to which states are open to new information about best practices (Gilardi, 2010). High student performance and a high proportion of highly qualified teachers may indicate that a state already has a successful strategies for evaluating and coaching teachers, making them less likely to adjust their policies in response to RTTT.

Second, we incorporate measures of partisanship into our model, given that politics can matter just as much as the specifics of a given policy arena; consequently. States with a given party in power are more likely to adopt policies in-line with their ideology (Volden, 2006). Therefore, we might expect that Republican led states to exhibit a slight preference for adopting teacher evaluation practices, given that teachers’ unions, traditionally a strong Democratic ally, tend to oppose teacher evaluation systems. We use the difference in the number of Democratic to Republican seats in each chamber of the legislature to account for partisan power.

Third, we account for geographical proximity, a factor known to influence the likelihood of policy diffusion, by including a binary measure of whether or not two given states are contiguous. Existing work shows that policies are more likely to diffuse within geographic regions, especially among contiguous states (Berry and Baybeck, 2005). This suggests that neighboring states should be more likely to adopt policies from one another than from non-contiguous states. Descriptive statistics are given in Table 3.

Table 3: State-Level Descriptive Statistics

| Statistic | N | Mean | St. Dev. | Min | Max |
|---|---|---|---|---|---|
| Ln(Student Population) | 408 | 13.3 | 1.1 | 11.1 | 15.7 |
| Ln(Teacher Population) | 408 | 10.6 | 1.0 | 8.6 | 12.8 |
| Math Proficiency | 408 | 64.8 | 16.5 | 18.8 | 97.0 |
| Reading Proficiency | 408 | 65.8 | 15.8 | 23.8 | 94.9 |
| Pct Qualified | 408 | 95.4 | 9.6 | 1.0 | 100.0 |
| Dem Margin House | 408 | −3.4 | 37.9 | −196 | 128 |
| Dem Margin Senate | 408 | −1.9 | 14.0 | −32 | 32 |

### 3.2.2 Results

As states' policies become more similar, evaluation prevalence increases but the gap between states, $y_{ijt}$, decreases. Our hypotheses considered whether states' adoption patterns corresponded to competition, during which many states adopted similar policies at once, or learning, where states tended to follow other states who won awards in prior RTTT rounds. We use the **lme4** package to estimate the multilevel diffusion model (Bates et al., 2014), with results provided in Table 4.

We find that these data are consistent with the learning hypothesis, as the negative and significant coefficient on *Awardee* indicates that the prevalence gap decreased, meaning that states became more similar when the sending state $j$ was an awardee in the same period that the receiver state $i$ was not. The magnitude of this relationship—a 1-percentage point increase in similarity—is substantively significant given the overall mean prevalence of approximately 3.5%.

Adjacency had a small positive relationship to similarity but was not statistically significant. This is consistent with recent scholarship showing that diffusion networks are characterized by more complex structures than geographic proximity (Desmarais et al., 2015). The receiver states' status with respect to the percentage of highly qualified teachers had a small negative impact on similarity. This suggests that states who already have highly qualified and effective teachers may see less of a need for evaluation reform. The Democratic seat margin also

Table 4: Policy Diffusion Estimates

| | |
|---|---|
| $\text{Awardee}_j$ | $-0.010^{***}$ |
| | (0.002) |
| $\text{Adjacent}_{ij}$ | $-0.0004$ |
| | (0.002) |
| Ln(Student Population) | 0.005 |
| | (0.015) |
| Ln(Teacher Population) | $-0.005$ |
| | (0.015) |
| Math Proficiency | $-0.00003$ |
| | (0.0001) |
| Reading Proficiency | $-0.00003$ |
| | (0.0001) |
| Pct Qualified | $0.0002^{***}$ |
| | (0.0001) |
| Dem Margin House | $-0.0002^{***}$ |
| | (0.00004) |
| Dem Margin Senate | $0.001^{***}$ |
| | (0.0001) |
| Intercept | $-0.032$ |
| | (0.072) |
| N | 20400 |

$^{***}p < .01$; $^{**}p < .05$; $^{*}p < .1$

had small but significant correlations with similarity. Additional seats in a house chamber corresponded to a higher likelihood of diffusion while additional seats in a senate chamber corresponded to a lower likelihood. As house chambers tend to have more members overall, this result may reflect its greater capacity to learn from others or to introduce legislation.

## 4 Conclusion

In summary, we use the case of teacher evaluation reform to demonstrate how topics can be used as outcomes to study policy diffusion. Using a corpus comprising bills introduced by states between 2009 and 2016, we show that states enacted reforms in response to the Race to the Top grant competition by adopting the language of states who had already received awards. For future directions, we consider ways to account for different stages of the adoption process (introduction, enrollment, engrossment, etc.), expanding the analysis to include semantic content to compare state policies, and ways to quantify uncertainty when aggregating topic prevalence across units in a diffusion model.

# References

American Institutes for Research (2014). Timeline of relevant educator effectiveness research and policy.

Bates, D., Maechler, M., Bolker, B. M., and Walker, S. C. (2014). **lme4**: Linear mixed-effects models using **Eigen** and s4. http://cran.r-project.org/package=lme4.

Berry, F. S. and Berry, W. D. (2014). *Innovation and diffusion models in policy research*, chapter 9. Boulder, CO: Westview Press.

Berry, W. D. and Baybeck, B. (2005). Using geographic information systems to study interstate competition. *American Political Science Review*, 99(4):505–519.

Bischof, J. M. and Airoldi, E. M. (2012). Summarizing topical content with word frequency and exclusivity. In *International Conference on Machine Learning*.

Cohen-Vogel, L. (2005). Federal role in teacher quality: "redefinition" or policy alignment? *Educational Policy*, 19(1):18–43.

Desmarais, B. A., Harden, J. J., and Boemke, F. J. (2015). Persistent policy pathways: inferring diffusion networks in the American states. *American Political Science Review*, 109(2):392–406.

Doherty, K. and Jacobs, S. (2013). State of the states 2013: Using evaluations of teacher effectiveness to inform policy and practice. Technical report, National Council on Teacher Quality.

Gelman, A. and Hill, J. (2007). *Data analysis using regression and multi-level/hierarchical models*. New York, NY: Cambridge University Press.

Gilardi, F. (2010). Who learns from what in policy diffusion processes? *American Journal of Political Science*, 54(3):650–666.

Gilardi, F. and Fuglister, K. (2008). Empirical modeling of policy diffusion in federal states: The dyadic approach. *Swiss Political Science Review*, 14:413–450.

Gilardi, F., Shipan, C. R., and Wueest, B. (2017). Policy diffusion: the issue-definition stage. Unpublished Manuscript.

Grimmer, J. and Stewart, B. M. (2013). Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political Analysis*, 21:267–297.

Hanushek, E. A. (2011). The economic value of higher teacher quality. *Economics of Education Review*, 30(3):466–479.

Heck, R. H. (2007). Examining the relationship between teacher quality as an organizational property of schools and students' achievement and growth rates. *Educational Administration Quarterly*, 43(4):399–432.

Lavigne, A. L. and Good, T. L. (2014). *Teacher and Student Evaluation: Moving Beyond the Failure of School Reform*. New York, NY: Routledge.

Meade, N. and Islam, T. (2006). Modelling and forecasting the diffusion of innovation – a 25-year review. *International Journal of Forecasting*, 22:519–545.

Nicholson-Crotty, S. (2009). The politics of diffusion: public policy in the American states. *Journal of Politics*, 71(1):192–205.

Orfield, G. and Lee, C. (2005). *Why segregation matters : poverty and educational inequality*. Cambridge, MA: Harvard University Civil Rights Project.

Quinn, K. M., Monroe, B. L., Colaresi, M., H, C. M., and Radev, D. R. (2010). How to analyze political attention with minimal assumptions and costs. *American Journal of Political Science*, 54(1):209–228.

Rivkin, S. G., Hanushek, E. A., and Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica*, 73(2):417–458.

Roberts, M. E., Stewart, B. M., and Airoldi, E. M. (2016). A model of text for experimentation in the social sciences. *Journal of the American Statistical Association*, 111(515):988–1003.

Roberts, M. E., Stewart, B. M., and Tingley, D. (2017). stm: R package for structural topic models. *Journal of Statistical Software*.

Rothstein, R. (2004). *Class and schools: using social, economic, and educational reform to close the Black-white achievement gap*. Economic Policy Institute. New York, NY: Columbia University.

Salamon, L. M. and Elliott, O. V. (2002). *The tools of government: a guide to the new governance*. New York, NY: Oxford University Press.

Shipan, C. R. and Volden, C. (2006). Bottom-up federalism: The diffusion of antismoking policies from U.S. cities to states. *American Journal of Political Science*, 50(4):825–843.

Shipan, C. R. and Volden, C. (2014). When the smoke clears: expertise, learning and policy diffusion. *Journal of Public Policy*, 34(3):357–387.

US Congress (2009). American recovery and reinvestment act. Section 14005-6, Title XIV, (Public Law 111-5).

US Department of Education (2013). Race to the top executive summary.

Volden, C. (2006). States as policy laboratories: emulating success in the Children's Health Insurance Program. *American Journal of Political Science*, 50(2):294–312.

Volden, C. (2015). Failures: diffusion, learning, and policy abandonment. *State Politics and Policy Quarterly*, 16(1):44–77.