2020

How much do you care about education? Exploring fluctuations of public interest in education issues among top national priorities in the U.S.

Dana Nehoran
University of the Pacific

HOW MUCH DO YOU CARE ABOUT EDUCATION? EXPLORING FLUCTUATIONS OF PUBLIC INTEREST IN EDUCATION ISSUES AMONG TOP NATIONAL PRIORITIES IN THE U.S.

By

Dana Nehoran

A Dissertation Submitted to the

Graduate School

In Partial Fulfillment of the

Requirements for the Degree of

DOCTOR OF EDUCATION

Benerd College
Educational and Organizational Leadership

University of the Pacific
Stockton, California

2020

HOW MUCH DO YOU CARE ABOUT EDUCATION? EXPLORING FLUCTUATIONS OF PUBLIC INTEREST IN EDUCATION ISSUES AMONG TOP NATIONAL PRIORITIES IN THE U.S.

By

Dana Nehoran

APPROVED BY

Dissertation Chair: Rachelle Kisst Hackett, Ph.D.

Committee Member: Delores McNair, Ed.D.

Committee Member: Justin Low, Ph.D.

Senior Associate Dean of Benerd College: Linda Webster, Ph.D.

HOW MUCH DO YOU CARE ABOUT EDUCATION? EXPLORING FLUCTUATIONS OF PUBLIC INTEREST IN EDUCATION ISSUES AMONG TOP NATIONAL PRIORITIES IN THE U.S.

By

Dana Nehoran

ACKNOWLEDGEMENTS

Leaving the home port and sailing at full speed ahead would not have been enough to arrive at the shores of a complete dissertation without all those who have helped me navigate this intricate research ship.

First and foremost, I would like to thank Dr. Rachelle Kisst Hackett, my dissertation committee chair, for helping me steer the ship at times of turbulent waters, when I thought that I would rather go get another ship. In addition, the skills she had taught me in quantitative research methods proved vital throughout this journey, and just as important was her constant challenging of the concepts in my dissertation, which made me research deeper and deeper to find better solutions.

I'd like to thank Dr. Delores McNair, member of the dissertation committee, who not only provided me with the knowledge of qualitative research that was essential to the ship construction, but her inspiration and motivation along the way served as the compass that guided my research through previously uncharted waters.

Deep gratitude is in order to Dr. Justin Low for his lectures on structural equation modeling (SEM) and latent variables, which were key concepts utilized in this dissertation, and for his advice and time spent as a dissertation committee member.

I'd like to thank my children. My son Nadav was the first to challenge many of my technical assumptions, helping me refine key points. My son Barak, my daughter Effie and my son Roy provided me with constant feedback on research ideas.

Lastly, my greatest appreciation goes to my husband, Yoed, who in addition to providing me with unconditional support and motivation during this long and strenuous sail, also offered me technical and strategical guidance that made this trip possible and even fun. He kept reminding me that a smooth sea never made a skilled sailor. Without him, I would not have embarked on this trip in the first place.

HOW MUCH DO YOU CARE ABOUT EDUCATION? EXPLORING FLUCTUATIONS OF PUBLIC INTEREST IN EDUCATION ISSUES AMONG TOP NATIONAL PRIORITIES IN THE U.S.

Abstract

By Dana Nehoran

University of the Pacific
2020

It is well known that a strong education system produces citizens who are more engaged in civil and social duties, with obvious benefits to society and the individuals. Policymakers who have the power to help improve the education system frequently rely on the news or the polls to better understand the issues involved, but these tools are often unable to answer customized questions on the public view with a large enough coverage.

Monitoring the American public interest in education over the years is not new. In fact, a number of national polling agencies have tracked education as part of their larger polls asking people to name the most burning issues facing the US. While these polls provide a fair indication of the changes in importance of education in the eyes of the public, they do not identify the factors which have historically been associated with the major fluctuations of such importance. Most importantly, these traditional national polls do not track public concern about specific subtopics within education.

This mixed methods study includes the creation of a software instrument with the objective of exploring the salience of education as a national priority over time and analyzing the possible factors associated with these fluctuations of interest. In addition to discovering the most prominent latent subtopics affecting education (such as *academic achievement*, *sexual assault* and *freedom of speech*), this study also seeks national-level issues that may have recently been associated with the largest declines.

The only source of data utilized is the text of tens of thousands of published news articles. Terms extracted from the text using natural language processing serve as the basis for automated qualitative analysis. As topics emerge from the data, the frequencies of the terms are utilized to associate the articles with the most relevant ones.

The analysis shows that public interest in education has declined the most during election times. It is also found that the areas that contributed the most during the largest surges of public interest in education from 2015 to 2020 were school budget, academic achievement gaps and mental health.

TABLE OF CONTENTS

LIST OF TABLES

Table

LIST OF FIGURES

Figure

LIST OF ABBREVIATIONS

atf adjusted term frequency. 10, 57–60, 66, 69

freq frequency of a term in a document. 57

IPOR inferential public opinion research. 16, 22, 28, 47, 52, 53, 60, 67, 76

MIP most important problem. 34, 41, 53, 54, 70, 98, 102

NFO need for orientation. 32, 33

PIEDI Popular Interest in Education Issues. 44, 48, 49, 51–53, 55, 56, 62, 65, 69, 70, 74–79, 81, 83, 84, 91, 94, 100–102

QCA qualitative content analysis. 16, 45, 48, 77

SES socioeconomic status. 26

TA thematic analysis. 45, 46, 49, 84, 85

CHAPTER 1: INTRODUCTION

With 3.7 million teachers, more than 56 million students in elementary, middle and high schools, and more than 19 million students in colleges, education is one of the most significant expenditures in America. Education is the largest category of state funding in the US (West et al., 2009). Despite the importance of education as an integral part of our democracy, the media coverage, as well as the public opinion, place education in a very low position, when compared to other issues of national interest (West et al., 2009). Through a thorough analysis of interest in national issues from 2015 to the present, this dissertation evaluates the factors associated with the largest fluctuations of interest in education as compared to other issues people care about.

The literature review and the conceptual framework both constitute the foundation on which this dissertation is based. The study starts with an explanation of public opinion as a concept and explores how the public opinion can be inferred from published articles. After covering the importance of public opinion among policymakers and the contribution a healthy and solid education system makes to the society, this study explores academic research on public opinion, public and media agenda, salience measures of influence and polls to emphasize the need for a new study that describes the factors associated with fluctuations of interest in education over the years.

This research is performed through four main stages which together help explore the fluctuations of public interest in education over time, and the factors involved in the largest fluctuations since 2015 (see: Figure 4: Dissertation research design & roadmap). The first stage (A) is aimed at the creation of a software instrument capable of collecting tens of thousands of news articles from selected newspapers and producing a time series depicting the public interest in education. In the second stage (B) the instrument created is utilized to show the fluctuations of interest in all issues in the US over time. During the third stage (C) a time series analysis is conducted and by sets of annotated related documents are

produced. Ultimately, the fourth stage (D) results in the human interpretation of the factors associated with fluctuations of interest in education over the years.

Background

Public opinion research involves the analysis of data collected from multiple individuals, as a generalization method to understand collective opinions and attitudes (Hodgkiss, 2017; Vu, 2014). Over the years, many researchers analyzed the media's news prioritization and filtering process. They describe the role of the gatekeepers, newspaper employees in charge of publishing decisions who determine what is published and what is not, based on newsworthiness (Hodgkiss, 2017; Serban, 2015; Strömbäck et al., 2012; Vu, 2014).

Traditionally, gatekeepers of the pre-internet era were subjective and highly biased, hinging their decisions on intuition, judgment and prior experience, and perceived to be disconnected from their audience (Vu, 2014). In today's world of online publishing and interactivity, gatekeepers have access to previously unavailable information: real-time news consumption data through dashboards, reports, online reader response, and more (Vu, 2014; Welbers et al., 2016). Consumer-driven newspapers depend on advertisement for revenue generation, and as Vu (2014) mentioned, the is a harsh competition for advertisement share, due to the high availability of multiple alternative channels, including TV news and social media. The financial implications emphasize the importance of readership metrics and justify the constant monitoring of readership traffic by journalists (Vu, 2014).

As reported by Vu (2014), the newsroom at *The Washington Post* has a screen displaying web metrics, and journalists at "The Post" receive daily emails with audience metrics. Furthermore, MacGregor (2007) added that mass media organizations such as *The Guardian*, *CNN* and *AOL* implemented tools that provide almost instantaneous reports of readership metrics. Such metrics include, in addition to the number of people reading a specific story, the time spent on it, how often the article was shared, and even the level of

reader engagement with its content, measured by the number of comments made by readers and by the frequency of *"like"* clicks (Vu, 2014). Journalists can tell the preferences of their audience in "unprecedented detail" (Welbers et al., 2016). In fact, using the number of clicks, number of unique visitors, and the time spent on the online version of news articles, today's gatekeepers are also able to tell whether they were successful in the anticipation of the level of interest in a specific topic (Tandoc, 2014; Welbers et al., 2016). Moreover, journalists constantly analyze the list of their stories ordered by popularity metrics to make decisions whether to "deepen and extend a story or to abandon further development" on its topic (MacGregor, 2007). Stories are grouped in categories by content, and these groups are tracked for "future publishing decisions" (MacGregor, 2007). Gatekeepers even use private online posts (e.g., social media or blogs) as a measurement for newsworthiness (Yamamoto et al., 2017), which further confirms that their decisions are a reflection of the public opinion. With the rise of social media, and the abundance of news sources, readers are not just news consumers, but active participants in news creation and distribution, helping spread the news and affecting its content (Vu, 2014).

The limited capacity of the press (Serban, 2015) and the ever increasing number of issues that people care about (McCombs & Zhu, 1995), together with the financial motivation of consumer-driven gatekeepers, leads to a decision making process in newsrooms that gives preference to topics meeting the readership interest (Blanchett Neheli, 2018; Tandoc, 2014; Vu, 2014; Wallace, 2018). Over the years, many scholars (Martin, 2013; McCombs, 2004; McCombs & Zhu, 1995; Tran, 2013) studied the relationship between the media agenda (the issues deemed as top priority by gatekeepers) and the public agenda (the issues deemed as top priority by the public), and found that they were highly correlated, meaning that the more salient a topic is in the news, the more important it is perceived by the public (and vise-versa).

Although scholars never agreed on a single definition for the term *public opinion*, what is clear to all of them is that people have opinions (Bardes & Oldendick, 2012; Glynn

et al., 2016c). As described by Glynn et al. (2016c), these can come in the form of direction (for example, a person may be in favor of the president's actions), intensity (for example, a person can be much more interested in education than in economy), or stability, referring to the degree to which an individual maintains their opinions over time.

The literature shows that extent of press coverage of a specific topic (called *object* in Communication Theory) is highly correlated to public interest (McCombs, 2004), and such interest can be inferred by the text contained in published news articles monitored by professional consumer-driven gatekeepers (Nehoran & Nehoran, 2020). The concept of inferential public opinion research (IPOR) was developed by Nehoran and Nehoran (2020) to guide the development of instruments capable of assessing the public interest in relation to a topic of interest. This dissertation uses the IPOR development model to create an instrument able to infer issues of public interest through a qualitative content analysis (QCA) of the text of news articles.

As shown by polls, academic publications and newspaper coverage, *education*, as an issue of national priority, has been historically displaced by many other issues Americans care about, such as economy or unemployment (Jackson, 2020; Jones & Saad, 2020). This dissertation will complement these studies by analyzing the factors which make the salience education decline or grow as an issue of national priority.

Problem Statement

The public generally believes that education is a national priority, but when it comes to prioritizing education among all other issues facing the US, the importance of education in the eyes of the public is considerably lower. In fact, less than 5% of the people mention the word *education* when they are asked what they think the top priority in the US is (Jackson, 2020; Jones & Saad, 2020). This either means that the public is more preoccupied with other issues, or that the public is not well aware of the issues of education due to lack of media exposure.

Based on the conceptual framework of this dissertation, a lack of interest in education may cause disinterest of policymakers, which affects the governmental support of the education system (Bardes & Oldendick, 2012). Politicians constantly check the public opinion to ensure that their actions meet the needs of their constituents (Burstein, 2003). If education is not a top priority of their voters, nor will education be a priority of the policymakers.

The factors that cause education interest to drop, in relation to other issues of national significance, are currently unknown to scholars, neither are the factors that increase the public interest in education.

This dissertation created an instrument able to explore the salience of education as a national priority over the years and analyze the possible factors causing fluctuations in interest. By understanding these factors, decision makers will be able to evaluate methods aimed at preserving high levels of public interest in education, which in turn will encourage policymakers to propose laws to improve education.

Conceptual Framework

The conceptual framework of this dissertation comprises four distinct components, which together encompass the ecosystem in which this study operates (see Figure 1). The framework starts with the concept that public opinion can be inferred from published news articles (Nehoran & Nehoran, 2020), and continues with the importance the results of public opinion have for policymakers working on education issues (Burstein, 2003). These legislators propose laws aimed at improving the social, economic, cultural or human types of capital produced by the education system (Bartee & Brown, 2007), which in turn make the citizens more engaged in the civil and social spheres of the society (Campbell, 2006).

First is the IPOR Development Model, which suggests the methodology for the creation of an instrument able to infer public opinion from the text of a large repository of documents maintained by professional consumer-driven gatekeepers (Nehoran & Nehoran, 2020). This development model constitutes the basis of this study, which collects news

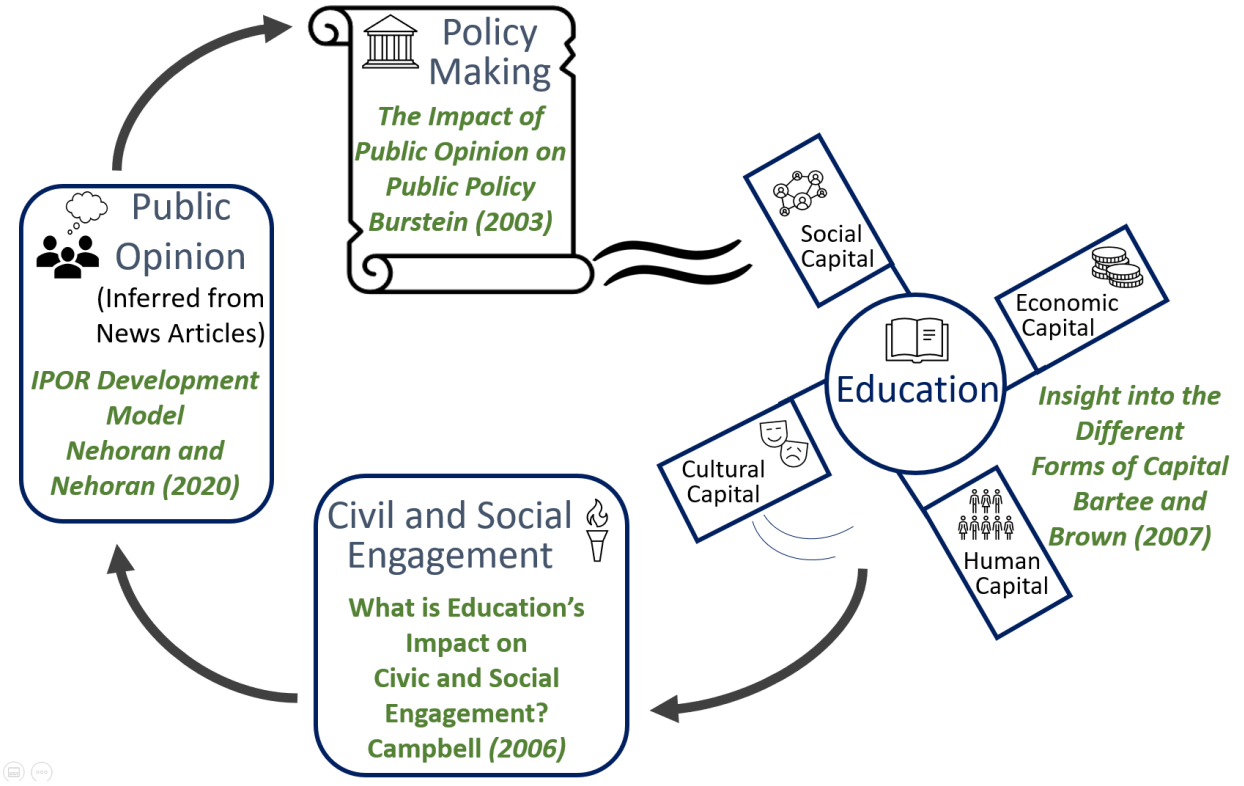


Figure 1. Conceptual framework

articles published online by newspapers recognized for archival by the Library of Congress. The second component is the publication named *The Impact of Public Opinion on Public Policy*, by Burstein (2003), containing empirical evidence showing the influence of public opinion on policy-making decisions. Third is the work by Bartee and Brown (2007), *Insight into the Different Forms of Capital*, which covers the relationship between the economic, human, cultural and social capitals as they interact within the education system. Lastly, the conceptual framework includes the publication by Campbell (2006) titled *What is Education's Impact on Civic and Social Engagement*, which emphasizes the impact that education has on civil and social engagement (CSE), which cyclically leads to more interest in education.

With the support of these four components, this study a) creates, validates and implements an IPOR-based instrument able to infer the level of public interest on

education issues over time and b) analyzes the possible factors that may have contributed to the largest fluctuations of such interest. Factors affecting the public interest in education impact the decisions made by policymakers, which affects the education system and its contribution to society.

Defining Public Opinion

"... public sentiment is everything. With public sentiment, nothing can fail. Without it nothing can succeed."
– Abraham Lincoln

Traditional democracies are built upon the assumption that the people are aware of issues of public interest and capable of making decisions that trigger actions (Bardes & Oldendick, 2012). In fact, Bardes and Oldendick (2012) describe the *"American mind"* as the views of adults on policy issues. Among the numerous occasions in which the public opinion shaped governmental decisions, one can mention the decision of Lyndon Johnson not to run for reelection, or of Richard Nixon to resign from the presidency, both influenced by the public disapproval (Bardes & Oldendick, 2012).

The definition of the term *public opinion* is a topic of historical debate among the scholarly community, in which their members hold different (and sometimes controversial) views of what it means, what it encompasses, or how it should be measured (Bardes & Oldendick, 2012; Glynn et al., 2016c). It is either defined as the preferences within a group of people, as the aggregation of individual thoughts about issues impacting communities, or the views of individuals over controversial issues affecting the society (Bardes & Oldendick, 2012). Even though scholars do not agree on a single definition of public opinion, there are some common elements in their descriptions that could be accounted for in an attempt to describe what public opinion is, but not before a clear understanding of what *public* means.

There are three social formations mentioned by scholars: *crowds*, *masses* and a *public*. Glynn et al. (2016c) defined *crowds* as a behavior of multitudes who act without personal responsibility, adopting ideas from the rest and behaving in ways accepted by the

group, but not necessarily conforming with the ways they would normally behave without the influence of the multitude. An example mentioned by Glynn et al. (2016c) is the effect cheerleaders have on the audience, inciting them to stand, yell, jump and perform other acts they would not normally do.

Crowds share motivations and experiences, while *masses* are basically groups of isolated, anonymous individuals who hold minimal (or no) interaction between them. Being detached from the rest of the people removes the need for conformity with the rest, allowing them to hold more individualized opinions (Glynn et al., 2016c). Masses and crowds have, among other examples, purchase power because their preferences may influence the type/quantity of products being produced or their prices.

Blumer (1953) defined a *public* as a group of people with divided ideas who engage in conversations about an issue, with the ability to think and empathize with other people in the public. This definition could potentially have distinguished between masses and a public (which actively engage in conversations over topics of interest). However, controversies arose with this definition, mainly with arguments citing the impossibility of engagement of the entire population of a country. Furthermore, since most Americans engage in some sort of participation on issues they care about, the definition of masses and a public are interchangeable for many scholars (Glynn et al., 2016c).

According to Glynn et al. (2016c), there are scholars who believe that *public opinion* is the aggregation of individual opinions, while others think that it is the opinion of the majority. Some argue that public opinion is expressed through interest groups, or by elite members of the society. Bardes and Oldendick (2012) defines public opinion as the *aggregate of the views of individual adults on matters of public interest.*

Dimensions of public opinion. Although scholars have different definitions of public opinion, they all agree that the public has opinions, but it is not always clear how to collect them (Glynn et al., 2016a). In fact, just a few questions asked in a poll cannot possibly reveal all of what people believe (Bourdieu, 1986; Glynn et al., 2016a).

Opinions can be categorized into three groups (see Figure 2): the ones that express *direction*, the ones that show *intensity*, and the opinions that represent *stability* (Glynn et al., 2016a).

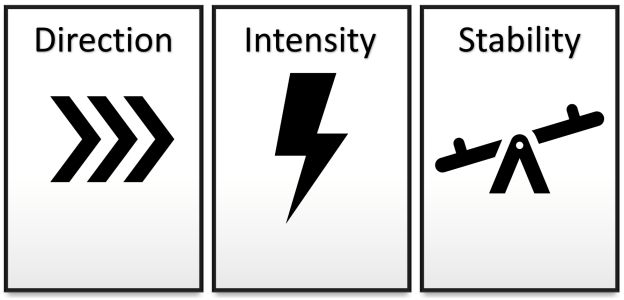


Figure 2. Dimensions of public opinion

The *direction* shows the attitude, or sentiment, towards a certain issue or topic. For example, some people may be in favor of gun control, while others may be against. The same applies to many other issues, like investment in education or abortion. The direction is not always clear, because there are occasions in which the answer is "depends": some people may be in favor of capital punishment in certain situations and against in other ones (Glynn et al., 2016a).

The *intensity* represents the interest level of the public, how strongly people care about an issue. Issues of high interest are usually deemed as "high profile" and attract attention from the press and policymakers. As pointed out by (Glynn et al., 2016a), a dilemma arises when policymakers have to decide whether to support high-intensity issues of the majority, or of the minority groups, when big differences show.

The *stability* of opinion denotes the consistency of opinions over a period of time. Stability is known to drag more attention from the leaders than the ones that are sporadic or that change with frequency (Glynn et al., 2016a).

How Public Opinion Can Be Inferred: Component One

Public opinion, expressed as the mood, sentiment, consumer preferences, or controversial ideas, can be collected and measured in a methodological way, in four main ways: survey research/polling, focus groups, experimental research and mass-media content analysis (Glynn et al., 2016b). When mass-media publications are monitored by professional, consumer-driven gatekeepers, the articles that are deemed as newsworthy (and hence published) are the ones that represent the readership of the newspaper (Hodgkiss, 2017; Serban, 2015; Strömbäck et al., 2012). In fact, the gatekeepers constantly monitor the performance of the news they publish to increase their profit, publishing more about topics with good audience response (measured by the clicks and time spent on published articles), and reducing the coverage on topics with poor readership interest (Hodgkiss, 2017; MacGregor, 2007; Welbers et al., 2016).

IPOR-based instruments analyze topics related to public opinion (Nehoran & Nehoran, 2020). These topics are created after the research questions of the study at-hand are formulated, and after a *benchmark survey* that answers a verifiable derivative research question is selected. Such a survey can be an existing one, or a new one conducted for the purpose of the study. Upon completing multiple cycles of quality assurance and calibration of a prototype of the software instrument, a report including the methodological choices made by the researcher is produced.

Public Opinion and Policy Making: Component Two

One of the promises of a democracy is that policymakers care about the opinions of their constituents, who believe that their views can be translated into public policy (Bardes & Oldendick, 2012). Despite the difficulty of linking the opinion of the public and laws being proposed, multiple methods examining the congruence between opinion and policy proved an existing correlation between them (Bardes & Oldendick, 2012). These methods show that even though not all public desires make it into policy, a significant number of them do. In fact, policymakers (at all levels) consult polls not only for decision making,

but to identify the problems of high importance to their constituents (Bardes & Oldendick, 2012).

The second component of this conceptual framework is the publication by Burstein (2003), which analyzed the impact public opinion had on policy making. Using a meta-analysis of multiple articles measuring the impact of public opinion on policies, Burstein analyzed weather public opinion affected public policy and if salience of issues increased such influence.

As reported by Burstein (2003), whether or not the public opinion matters has been a topic of controversies over the years, with some scholars claiming that governmental representatives pay minimal attention to the public, while others stating that the opinion of the citizens was highly regarded by policymakers. Although these two theories seem extremely opposite, they both agree that the opinion of the public is considered, but they disagree as to the extent in which it is. In his research, Burstein concluded that public opinion had a significant effect on policy making *"most of the time, often strongly"* (Burstein, 2003, p. 29).

An additional objective of Burstein's research was to evaluate if there was a correlation between the salience of issues and government responses to them. In other words, is responsiveness to issues higher as they are more talked about by the public? His findings showed that the salience of issues affected the influence exercised by public opinion on policy making. In general, scholars believe that a) competition over electoral votes makes political parties pay attention to the public opinion, b) the role of interest groups is to keep legislators informed about the opinions of the public, and c) the public opinion is relevant even when political parties or elite organizations hold ideas that are contrary to the public beliefs (Burstein, 2003).

In summary, public opinion counts, either through the political organizations or elite groups, which act as liaisons between the people they represent and the legislators, or through the electoral power of masses, which may determine which candidate will be

elected, recalled, or replaced. The more salient the issues of public interest are, the more attention they get among policymakers. Public opinion matters, even when political parties or elite groups oppose the views of the public. The term *improvement* is a relative noun, particularly when it relates to legislation, due to the multiple political views held by the public. What is deemed as beneficial for one policymaker (e.g. the incorporation of the common core in schools), may be thought as harmful by others. However, what is clear is that policymakers make decision in an attempt to make improvements. When lawmakers take the initiative and propose laws to improve education, they not only make an impact on the individuals attending school, but they contribute to the four forms of capital (initially defined by Bourdieu (1986)), as explained in the following section.

Education Contribution to Capital: Component Three

The notion of capital in its four forms (social, cultural, human and economic) was introduced by Bourdieu (1986), as accumulated assets that can produce profit and expand over time. Capital, which can be non-economic (e.g. cultural capital), may be acquired, exchanged and converted from one form to another. Bourdieu (1986) posited that a deep understanding of the different forms of capital would help explain the functioning of the social world.

The third component of this conceptual framework centers on the work of Bartee and Brown (2007), which explores the interaction between all forms of capital in the education context and explains in detail each one of these forms. Economic capital includes all forms of financial assets, including, but not limited to, financial currency, individual possessions and all goods that can be sold, exchanged, or bartered. Human capital is the accumulation of knowledge, ability or skills, and is often measured as the likelihood of return of the investment placed on education.

Cultural capital is associated with ethnic customs, traditions and beliefs that provide individuals with a cultural identity, knowledge, values, and dispositions. Lastly, social capital refers to the accumulation of social connections through influential networks,

associations, or affiliations to institutions, all made to gain access to material and non-material assets.

The association of economic capital with education was originally explored by Bourdieu (1986), in the context of its reproductive aspects. The general notion that prevailed then was that schooling systems (originally set up with the norms of elite classes), performed as systems of reproduction of social classes. In today's world, financial assets from families utilized to support student education are included in the classification of economic capital (Bartee & Brown, 2007). In addition to families, education institutions, too, play a role in economic capital by owning tools, equipment and buildings, acquired or built with the objective of the creation of additional capital through education (Bartee & Brown, 2007). Students, as well as education institutions, act as both consumers and producers of economic capital. Students' investment of financial assets on tuition, and time that would otherwise be spent on lucrative activities, all contribute economic capital of the education institutions, which in turn helps individuals gain economic capital through increased financial prospects. Society, on the other hand, contributes financial assets in the form of taxes that subsidize the costs of education and receives economic development to support social services through trained and skilled citizens (Bartee & Brown, 2007).

The next form is human capital, which encompasses the acquired abilities and capabilities that enhance the freedom of choice and productivity. These include knowledge, attitudes and skills that can be applied to gain access to new opportunities. Meritocracy is often mentioned alongside human capital, as a way of expressing the value produced by the education system, which helps individuals utilize their gained human capital into economic capital through increased earnings resulting from a degree (Bartee & Brown, 2007).

Cultural capital can be tangible, such as artworks and cultural heritage, or intangible, such as ideas, beliefs and practices. Scholars disagree on the impact cultural capital has on educational attainment. While Sullivan (2001) agreed that cultural capital contributed to school retention, she explained that when cultural capital was controlled for,

social class affected educational attainment, meaning that cultural capital was a partial factor that increased retention. On the other hand, Wells (2008) concludes that cultural capital contributes to persistence among first- and second-year college students. In addition, Jæger (2010) found that (with differences between SES groups) cultural capital had an influence on reading and math scores. Education institutions are key contributors of cultural capital, through cultural activities, arts workshops, multicultural education and parent involvement in education activities (Bartee & Brown, 2007).

Social capital is the interpersonal communication and exchange among members of a society, including networks, relationships and mutual acquaintance (Bourdieu, 1986). Education institutions encourage the development of social capital through organized activities for faculty, staff and students and create spaces for scholarly exchange, which cultivates the social capital of the individuals involved (Bourdieu, 1986).

Educated individuals are more likely to read the news (Popkin, 1991), which should make them more informed about the societal issues impacting their community, and hence more likely to contribute to the civil and social aspects of the community they live in, as explained in the next section.

Education and Civil and Social Engagement (CSE): Component Four

> *"My hypothesis is that education affects politics not by deepening but by broadening the electorate-by increasing the number of issues that citizens see as politically relevant, and by increasing the number of connections they make between their own lives and national and international events"*
> – Popkin, 1991

The last component of this conceptual framework is based on the article *What is education's impact on civil and social engagement?* by Campbell (2006), who introduces his paper explaining the importance of education as a producer of social capital. His publication defines seven dimensions of engagement cultivated by the education system: political engagement, civil engagement, voting, trust, tolerance, and political knowledge. Campbell (2006) makes a distinction between *political engagement*, which involves influence

efforts by the citizens, and *civic engagement*, which is passive participation with no desire to affect public policy. On the other hand, according to him, *voting* belongs to its own category of engagement. *Trust*, consisting of a mindset of belief, along with *tolerance*, build a solid society of collaboration and cooperation (Campbell, 2006). Lastly *political knowledge* empowers citizens with knowledge about democratic institutions, which helps them make quality decisions that meet their ideals.

The study by Campbell (2006) revealed that, while increased levels of education did not boost citizenship (political and civil) engagement, post-secondary education had a causal relationship to voter turnout, political tolerance, interpersonal trust, group memberships and the acquisition of political knowledge (mainly through newspaper reading). The level of education does not seem to influence the time spent watching TV, but it considerably improves the probability of that individuals will read newspapers (Popkin, 1991). Not only that educated citizens read the news considerably more than non-educated ones, but they discuss news stories more often that the rest (Popkin, 1991). Higher levels of education, together with this increase in newspaper reading, has a positive correlation with the number of issues of national interest individuals can recognize (McCombs & Zhu, 1995). Campbell (2006) argued that if education institutions increased bureaucratic competence, reinforced the civil and cognitive skills of their students and provided a safe environment to debate political issues in the classroom, the level of civil and social engagements of the students would grow and their contribution to the society would increase accordingly.

Summary of the Conceptual Framework

This study is based on a conceptual framework consisting of four components that constitute the foundation upon which the methodology and importance of this dissertation is built upon. The data collection and analysis stages of this dissertation are based on the first component, the *IPOR Development Model*, by Nehoran and Nehoran (2020), which explains the process by which the public opinion can be inferred from published articles

maintained by professional consumer-driven gatekeepers. The second component, *The impact of Public opinion on Public Policy*, by Burstein (2003) establishes the foundation for a deep understanding of the importance that public opinion on policy decisions, as policymakers constantly monitor public opinion to ensure their actions meet their constituents' wishes. The more people care about education, the more laws aimed at improving education are passed. The third component *Insight into the Different Forms of Capital* by Bartee and Brown (2007) covers the impact education has on the four forms of capital, and the fourth component *What is Education's Impact on Civil and Social Engagement*, by Campbell (2006) explains the contribution education has to the civil and social engagement of citizens.

In summary, when education becomes important for the public, it draws the attention of policymakers, who take actions to improve the education system, which in turn produces more educated citizens.

Significance

Public opinion on education issues matters because policy makers and interest groups influencing policy makers constantly analyze polls to monitor constituents' sentiments and preferences. The more salient these issues are (measured by the importance given by the citizens), the more attention they get by governmental entities. When education issues get the attention of law makers, they may translate into laws aimed at improving the education system, which in turn positively increases the social, cultural, economic and human forms of capital. More educated individuals are more engaged in civil and social duties and are likely to help improve the rules by which the society operates.

The novel IPOR-based instrument developed for this study was able to identify the major fluctuations of interest in education and describe the factors associated with these fluctuations. In addition to exploring the national issues competing for public attention during education interest plunges, this study could describe which education issues caused the largest surges of interest in education.

Research Questions

This study explores the interest of the public in education issues over time by analyzing news articles monitored by professional consumer-driven gatekeepers.

The research questions are:

1. How has the interest in education, among all the national priority issues, changed in the US from 2015 to 2020?
2. When were the largest fluctuations in the public's interest in education?
3. Observing the steepest declines, which national issues displaced the public interest in education?
4. What were the specific education issues that caused the interest in education to surge?

Closing Remarks

Polls are the most popular measurement tools, and are widely utilized by governmental officials, political candidates, public and private institutions, corporations, and others. Polls are instruments able to collect information in a methodological way to answer a question of interest.

If the level of public interest in education increases, policymakers will be encouraged to pay attention to the issues impacting education. More educated individuals become involved citizens that are more able to contribute to society.

The next chapter explores what makes users read the news and the intertwined system of influence that exists between the media and the public. The more salient an issue is in public eyes, the larger the demand for news articles covering it is. This growing demand influences newspaper's gatekeepers to increase the salience of articles related to that topic. After explaining how salience is measured and how influence is defined, the chapter covers the grounds for a correlation between the media agenda (top priority issues among gatekeepers) and the public agenda (top priority issues perceived by the public). Despite the importance of education in our society, as described in the conceptual framework of this study, the literature review shows that education does not draw enough

attention in the eyes of the public. Lastly, a better understanding of the factors that impact importance of education in public eyes will help find ways of increasing the perceived value of education to ultimately improve to our society.

CHAPTER 2: LITERATURE REVIEW

In your opinion, what is the most pressing economic issue today? How would you answer that question? Unless you, or someone in your close social network, is involved in economic affairs, you would not be able to answer that question promptly and without hesitation. The same would happen if the question were about crime, since most of the citizens do not have firsthand exposure to it, or about immigration, just to mention a few. It is very likely that in order to answer the question you will seek information in the news - nowadays likely online.

Searching for news articles related to your topic using a web search engine sends a signal to the engine provider reporting your interest for such topic (Tandoc, 2014), which is later used to inform decision makers of many sorts. In addition, once you click the news article of choice, the news outlet that published that article can track the time you spent in the article, your navigation patterns in their website, and advertisement convergence, which is ultimately translated to profit for the news outlet (Christin, 2018; Tandoc, 2014; Vu, 2014).

In consumer-driven newspapers, increased traffic leads to financial gain. Therefore, gatekeepers monitor readership interest to increase the exposure of topics of popular interest and decrease the coverage of the ones that dwindle in popularity (Vu, 2014). In fact, your interest (or disinterest) level for a particular topic, aggregated to the interests of millions of people, influences the decisions of gatekeepers to increase, decrease or even discontinue the coverage that topic.

The technology of today allows gatekeepers to access real-time metrics of readership interest and act upon them. Therefore, an increase of news coverage of a specific topic triggers a corresponding increase of public interest on that topic, and vice versa (Christin, 2018; Hodgkiss, 2017; MacGregor, 2007; Serban, 2015; Strömbäck et al., 2012; Wallace, 2018; Yamamoto et al., 2017).

Literature Review Roadmap

The literature review starts by covering the motives that make individuals read the news. People feel uncomfortable when they do not have all the knowledge needed to make decisions, which often drives them to seek for information in the news - likely on the web. The review continues with a measure of *salience* of an item (or issue), which can be measured, and refers to the importance given to the issue by the media or by the citizens. The media can intensify the salience of an issue by increasing the frequency of coverage of that issue or by making the issue more prominent in the news. The salience of each issue covered in the news is determined by the gatekeepers, who constantly monitor readership dashboards to ensure maximum revenue from publications. There are mainly three players involved in issue salience: the media (which may be influenced by advertisers, wealthy individuals or corporations), the public and the public officials. These three players constantly increase or decrease the salience of issues based on internal or external factors. In Communication Theory, *influence* occurs when the salience is transferred from one player to the other one. For example, when a topic (e.g. *homelessness*) becomes more prevalent in the news, it is likely to affect the interest of the readers, hence influence them to be willing to read more articles related to that topic. After covering the degree to which the media influences the public opinion, and vise-versa, the review analyzes the coverage education issues have had in the media through the years. The literature review concludes with a thorough analysis of education salience in written news and in the polls, and evaluates the significance attributed to education against all other issues of national priority in the US.

What Makes Individuals Read the News?

People have a general need for closure - they are impatient and feel uncomfortable being in the information-seeking status (McCombs, 1968). Humans have an intrinsic necessity to organize their thoughts and arrive to a convincing conclusion that meets their ideologies, a concept denoted *need for orientation (NFO)*, referring to the act of searching for answers in the news (McCombs, 1968; O'Keefe, 1978; Shaw & McCombs, 1977). NFO

as a concept was first introduced by McCombs and Shaw (1972), and followed by many other scholars, encompassing the need of individuals to understand the world, particularly about issues not in their area of expertise (McCombs, 1968; McCombs et al., 2014). The need for orientation was explained as the psychological stage of discomfort with unfamiliar situations in which individuals do not have the knowledge they believe they need (Matthes, 2005). To determine the level of interest an individual has on reading the news on particular topic, scholars determined two factors: uncertainty and relevance (Matthes, 2005; McCombs, 1968). The individual's interest on reading a news article is highly dependent on how relevant the topic of the article is and how uncertain the person is about the topic (Matthes, 2005).

NFO's psychological effect drives individuals to search for information, most likely on the web. Is it possible that the simple act of reading the news will make the person has a strong opinion on a topic previously unknown? Can the news make individuals change preconceived opinions or sentiments? Furthermore, how likely it is that a person with a strong established opinion will change their mind after reading the news? It is all about influence...

Salience: How Can Influence Be Measured?

In Communication Theory, the term *object* denotes an item of interest of the press and polls (McCombs, 2004). Although the most frequently referred objects are issues and political figures, many other objects gained popularity over the years, such as goods, brands, institutions (e.g. educational), or corporations (McCombs, 2004).

Objects do not come in silos. For example, it would be difficult to find a published news article about a president that does not mention some of their characteristics or a few descriptions of their actions. The same would apply to any other object. For example, it would not make sense to publish an article about student homelessness without some explanation about aspects related to the issue. These characteristics, or descriptors, are called *attributes* (McCombs, 2004). The measure of *salience*, the most common one in Mass

Communication and Political Science fields, represents both prominence and perceived importance of a topic (Lippmann, 1922; McCombs, 2004; Tran, 2013). Salience can be measured from the media point of view (*media salience*) or from the public perception (*public salience*), as described by Tran (2013) and McCombs (2005).

Printed media salience has three dimensions: *attention*, depicting the frequency; *prominence* specified by the placement, text size and length; and *valence*, determined by the tone or level of emotion attributed to the object or its attribute (Tran, 2013). For example, if an article about a certain topic is placed on the landing page (the main page) of a website, the gatekeeper is using the prominence dimension of salience to increase traffic to that topic (or article).

If multiple articles of the same topic are published with high frequency, the gatekeeper is attempting to increase the attention of that specific topic. If the title of the article contains sensational text, the gatekeeper is trying to draw readership attention towards that article using the valence dimension of salience. Only a handful of articles can occupy a web landing page (Groshek & Groshek, 2013), and only a few of them can have more prominent features than the rest (e.g. enlarging the size of the title of all articles in the main page will not emphasize any of them). It is the gatekeeper's role to determine which articles to include and what will be the salience of each of them (Vu, 2014; Wallace, 2018). How is this decision made? Gatekeepers have multiple sources that they consult, one of which is a dashboard of readership interest and monetary conversion (through advertisement), which shows which topics have been popular and deserve increase coverage, and which ones do not (Christin, 2018; Tandoc, 2014).

On the other hand, public salience, which represents importance in the eyes of the public, is commonly measured by polls, hits on news or even Google trends (Tran, 2013). Measuring public opinion through dashboards showing the online behavior of the readership is unobtrusive and unbiased (Tran, 2013). The most common method of public opinion collection is the MIP (Most Important Problem) survey, which is administrated

regularly by the major survey agencies (Kim, 2013; McCombs, 2004; McCombs et al., 2014; Tran, 2013; Wanta & Alkazemi, 2017), such as Gallup or Associated Press.

Influence is defined as the transfer of salience from one player to another one. Wanta and Alkazemi (2017) described three actors of the agenda-building process: the news media, the public and public officials (see Figure 3). The agenda-setting theory, first introduced by McCombs and Shaw (1972), concentrates on the one-way influence exercised from the press to the public. This theory analyzes if a variation in media salience causes a correlated variation in public salience.

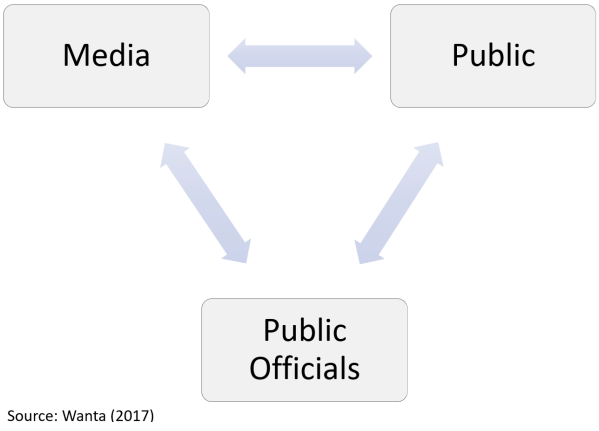


Figure 3. Intermedia agenda-setting

A relatively new theory in Communication Theory is the *intermedia agenda-setting theory* (Groshek & Groshek, 2013; Stern et al., 2020; Wanta & Alkazemi, 2017), which recognizes the media as being both influencers and influenced. The theory claims an interrelationship between these three actors (Wanta & Alkazemi, 2017), in a way that objects are part of a cycle in which the media, the public and public officials influence each other to conform with economic and social norms (Stern et al., 2020).

Increasing (or decreasing) the level of salience of an object (e.g. a political figure) can be done by impacting the exposure of the object, for example by increasing the frequency of articles related to that person. Concentrating on the object is not the only

way of affecting the level of salience. For example, a news outlet may increase the salience of a specific attribute of a politician by emphasizing more certain political tendencies of theirs. Furthermore, it is also possible to emphasize the association of characteristics of the political figure. McCombs (2004) defined three levels of salience to address the different kinds exposed here. The first level is concentrated on the object, guiding the user on what to think about (for example *Harvard*). The second level highlights the characteristics of the object (for example *admission criteria*). The third level of influence bundles concepts together by making the set of associated characteristics of the object more salient (for example *admission criteria, legacy admissions, bribe, discrimination, SAT, equal opportunity*). Affection (e.g. positive or negative sentiment) is part of the second level of influence. In fact, the news may influence the sentiment toward a political candidate, a brand or even an ethnic group (for example, approval of the president, or sentiment towards a new movie).

To What Degree Does the Media Influence Public Opinion?

The agenda-setting theory was first introduced by McCombs and Shaw (1972). It was inspired by the book by Lippmann (1922), and its chapter titled *"The World Outside and the Pictures in our Heads"*. Lippmann (1922) stated that people reacted to their perception of the world, based on what is exposed to them in the news media, rather than the real world. Since then, a plethora of articles were published by researchers, such as Wanta and Alkazemi (2017), Matthes (2005), Shaw and McCombs (1977), showing the influence of the press on public opinion.

Most citizens do not have previous knowledge about the functioning of the government, or about the most current national/international issues, and resort to the news to get informed (McCombs & Shaw, 1972). When an issue becomes important, it gets more coverage in the news (McCombs, 2004; McCombs & Shaw, 1972). Candidates use the media to gain exposure to people and influence their voting decisions (McCombs & Shaw, 1972).

People that are more educated are more politically involved and, thus, are more likely to read the news in seek for information (McCombs & Shaw, 1972). When the mass media present a topic, the public analyses it and shares it with others, which contributes to the salience of that topic in the public agenda (McCombs, 2004). These studies and multiple other agenda-setting ones show strong evidence that the media influence the public opinion. How about the public opinion, does it influence the media?

To What Degree Does Public Opinion Influence the Media?

Since the media and the public are two actors in the intermedia agenda-setting theory, they both play dual roles, as influencers and influencing agents. News in the media affect the public using any of the 3 levels of influence described in the agenda-setting theory McCombs and Shaw (1972). At the same time, the gatekeepers, who make publishing salience decisions, constantly consult dashboards to ensure maximum profit by increasing salience of popular topics and decreasing the ones with declining readership (Christin, 2018). The newsrooms include big screens that display metrics of many sorts, among them web analytics measures which allow the tracking of online traffic, such as clicks, time spent on site or advertisement conversion (Tandoc, 2014). Reader metrics ensure that the news published is relevant and meet the readership interest (Vu, 2014).

Does News Content Correlate with Public Opinion?

Media influences the public opinion, and public opinion influences the media, which means that there is a high correlation between the media and the public opinion (Martin, 2013; McCombs, 2004; McCombs & Shaw, 1972; McCombs et al., 2014; Tran, 2013).

In the Chapel Hill study by McCombs and Shaw (1972), the first of its kind, they found that the *"correlation between the major item emphasis on the main campaign issues carried by the media and voters' independent judgments of what were the important issues was +.967"*. A few years later, Weaver (1977) reported a correlation of 0.63 between the agenda of the TV and the public agenda during the primaries for the presidential campaign

in the US in 1976. A meta-analysis conducted by Wanta and Ghanem (2000), in which the authors analyzed close to a hundred agenda-setting studies, reported an average of 0.55 correlation between the media agenda and the public agenda (McCombs, 2004). Hundreds of empirical studies that reported high correlation were conducted around the world analyzing both, television and newspapers in all continents (McCombs, 2004).

Thanks to these studies that show a significant correlation between the media and the public opinion (regardless of which actor is the influencer), the public opinion salience can be used to infer media salience, and vise-versa. In other words, since media salience is highly correlated to public opinion salience, each can be evaluated interchangeably to infer the other one.

From "What Are the Most Important Issues?" Poll to Education

A quarter of a century ago, McCombs and Zhu (1995) conducted a study to quantify the results of the most popular open-ended question of all times: *"What do you think is the most important problem facing the country today?"* using Gallup's survey. The research sought quantitative answers to questions related to a) issue-carrying capacity (the number of issues nominated per person), b) issue variability (how varied are the issues mentioned) and c) issue volatility (how long, on average, issues stay in people's mind). In the words of McCombs and Zhu (1995, p. 517): *"The public agenda has been transformed from an era where one or two overriding issues dominated to the current stage where many voices compete for attention"*. They found that education was highly correlated with the variability and volatility effects. Their report shows evidence that more educated individuals read more newspapers, which contributes to a larger variety of issues of interest (variability).

As mentioned by McCombs and Zhu (1995), theirs was not the first study linking between public opinion and education. In fact, Popkin (1991) found that education affected the number of issues that individuals regarded as important and increased the associations they made between their own experiences and the world around them. Even

before that, Wade and Schramm (1969) analyzed newspaper reading behavior through surveys from 1952 and 1964, with survey questions along the lines of *"How much did you read..."*. They concluded that more educated groups were more likely to read the newspapers and magazines than the rest of the citizens.

The results by Popkin (1991) and Wade and Schramm (1969), together with the results from many other researchers, report that citizens (regardless of their level of education) typically do not have an upfront knowledge of issues of civil priority. However, people that with higher levels of education read the newspapers more, which increases their awareness and understanding of the events around world. This finding is in agreement with the concept of the framework of this dissertation, which emphasizes the contribution of education to the civil and social engagement of the citizens.

Education Coverage Is Invisible, or Is it?

> *"At the national level, we found that education coverage is virtually invisible. Except for specialized publications such as The Chronicle of Higher Education and Education Week, there is a dearth of education coverage in the mainstream media. Other than coverage of school sports, disease outbreaks, and periodic student crime sprees, the volume of coverage is certainly not equal to education's importance as a policy issue. Our review found that education simply does not generate the volume of coverage received by other major issues."*
>
> – West et al., 2009

Education, including elementary, secondary and tertiary education, is the largest category of state funding in the United States (Stauffer et al., 2017), with current expenditure per student of $\$13,440$, and the total expenditure of \$680 billion (*Back to school statistics*, 2019). Despite the 56.6 million students attending elementary, middle and high schools, and the 19.9 million students attending colleges and universities in America (*Back to school statistics*, 2019), *education* does not appear to be perceived as a priority by the U.S. citizens. In fact, less than 5% think about education when they have to determine what the top issues facing the US is (Hynds, 1981; West et al., 2009).

Education Salience Through Reporters' Perception

Academic research that analyzes education coverage in the US is not vast but started many decades ago. One example is the publication by Jacobson (1973). The research evaluated ways of improving education-news coverage by surveying a purposeful sample of education editors and eliciting information about attitudes towards education writing. Hynds (1981) wrote about the disproportion of coverage that education issues had in the agenda of the media. He analyzed the number of educational assignments dedicated to covering education at the metropolitan and national levels.

Similar to the research conducted by Jacobson (1973), the study by Hynds (1981) analyzed the factors that make the communication between the media and educators difficult. The findings revealed that education got more space than religion and arts, but less than other popular topics (such as sport, family and editorial pages). While coverage of education and business appeared in similar ratio, some smaller newspapers favored education to business (Hynds, 1981). The research by Hynds (1981) found that the topics covered included education controversies, board meetings and sport. Some other subjects covered were programs for gifted and disadvantaged students, drugs, school discipline, and inservice programs for teachers. The newspapers in the 80's used staff members, polls and syndicates as their sources of information (Hynds, 1981). The research by Hynds (1981) concluded that many of the newspapers had slightly increased their coverage of education and improved the quality of the content, by adding more interpretation and analysis to the reporting of education news.

The turn of the century was accompanied with the introduction of new technologies and norms of communication, such as blogs, YouTube videos, Twitter, Facebook, and the famous *like* button (West et al., 2009). While school systems keep a close communication with parents and blogs are used to debate education issues, these methods do not replace the systematic and traditional method of mass media (West et al., 2009). Local newspapers with journalists involved in school board meetings keep close contact with

school teachers and administrators, and are more likely to cover education topics than national outlets (West et al., 2009).

Education Salience Through Content Analysis of Written News

A study conducted by West et al. (2009) evaluated the coverage of education by media outlets. They analyzed hundreds of education stories published from 2007 to 2009 by local and national newspapers, Associated Press stories, and local education blogs. The study found that only 1.4% of the coverage in 2009 was related to education, very low compared to the 5% reported by West et al. (2009), a fact that incited the authors to denote the education coverage *"is virtually invisible"*. According to their report, the most common topics covered in the news in 2009 were government, economics, foreign affairs, health care, business and crime, in part because of pressing public issues, such as unemployment and a declining economy (West et al., 2009).

Lastly, West et al. (2009) reported that a very low percentage of education news focused on the quality of education, such as curriculum improvement, the process of learning, or reforms proposed for improvements. The news rather concentrated on episodes, crime, scandals and financial issues.

Education Salience Through Polls

With some variations in the text used, the MIP question was historically asked by pollsters for many decades. Two of the most renown agencies including the question at regular periods of time are Gallup and Ipsos (on behalf of Thomson Reuters).

The question asked by Gallup is *"What do you think is the most important problem facing the country today?"* (Jones & Saad, 2020), while the one from Ipsos is *"In your opinion, what is the most important problem facing the US today?"* (Jackson, 2020). Gallup's question is open ended while the one from Ipsos provides a list of choices with a note *"Select from below or write in"*. This distinction does not seem to matter too much at the first look, but providing the respondent with a list of alternatives prior to an

open-ended box may frame their train of thought and make a difference in people's reasoning to an almost identical question.

The options given to Reuters/Ipsos' respondents are *The Economy*, *Unemployment/lack of jobs*, *War/foreign conflicts*, *Immigration*,*Terrorism Healthcare*, *Energy Issues*, *Morality*, *Education*, *Crime*, *The Environment*, *Other*, *Don't know*. Another detail worth mentioning is that the question, in both surveys, asks for the single most important issue, which should discourage respondents from listing multiple important issues. However, some of them do include more than one issue in their responses, a fact accounted for by both agencies.

Both agencies interviewed adults living in all U.S. states in July 2020. Gallup used telephone and cellular phone interviews with a random sample of $1,007$ people while Ipsos conducted an online survey of $1,115$ individuals. Gallup reported that 1% of the participants mentioned education as the most important problem, while Ipsos survey resulted in 4.2% of the people mentioning education as the most important problem in the US. Since Gallup's survey was open ended, a large number of categories emerged from the text, some of which would have otherwise been aggregated into the education category. An example of such is the *school shootings*, which represented a category of its own.

There are many factors that may explain the difference between the results of Gallup (1%) and Ipsos (4.2%), among them are the distinction between the open-ended style of the Gallup's question against the hybrid style of Ipsos' question. Additionally, their target population of them differs because Gallup conducted phone interviews, while Ipsos conducted an online survey. Either way, both results correspond with the other studies analyzed in this dissertation, which report that the percentage of individuals regarding education as their top priority ranges between 1% and 5%.

Closing Remarks

This literature review started with a question about the most pressing economic issue today. What would you do if the question were about education? It is likely that you

would have conducted an online search about education in the US or about a specific topic of interest, such as *substance abuse in colleges*. Your search would have increased the education footprint on Google Search which, when aggregated with many other users inform decision makers that there was an increase in interest in education topics. Once you click on a few articles in the New York Times website, the gatekeepers at that newspaper can see the traffic increase in education. Since more traffic is translated to more revenue, gatekeepers keep monitoring the interest in education, and after a certain threshold they decide to increase the salience of articles related to education (by either increasing their frequency or their prominence). Such increase in coverage of education issues brings more exposure, which draws the attention of more readers, who react and read more education news. At the same time, more education stories attract policymakers' interest, who are constantly look for ways to please their audience, so they get reelected. New regulations are proposed by these policymakers, targeting the education system which, as detailed in the conceptual framework, plays a key role in our democracy.

Based on the analysis conducted in this literature review, a) there is a significant correlation between the public agenda and the media agenda, and b) education is barely covered in the news. Does this mean that people do not care (enough) about education? Consumer-driven gatekeepers have not historically assigned more than 5% of their agenda to education, meaning that there are other issues of national priority that people care about more.

CHAPTER 3: METHODS

The objective of this study was to analyze the largest fluctuations of interest in education issues over the years, and to understand the factors that influenced these fluctuations. Newspaper coverage as well as public opinion polls track interest in education over time, making it possible to compare it against other issues of national priority. A symbiotic relationship between the media and public opinion cause a mutual influence between them. When issues become prominent in public eyes, they influence news gatekeepers, who increase the salience of these issues in upcoming publications. Similarly, increased media coverage of a specific topic strengthens the related public awareness and the public interest in the issue. This high correlation between the media and public opinion allows researchers to study one of them and infer the second one.

The methodological foundation of this dissertation was based upon the first component of the conceptual framework, the *IPOR Instrument Development Framework* (Nehoran & Nehoran, 2020), which was used in this study to create a brand new instrument, *Popular Interest in Education Issues (PIEDI)*, to help infer the public interest in education over the years. This instrument is used in answering the four questions of this dissertation. Once created, it can be applied repeatedly, as long as the new sample fits the criteria of the instrument.

Approach

The technological evolution of computational tools brought an emerging trend to automate many of the labor-intensive tasks traditionally performed by humans, and qualitative research is not an exception. Qualitative analysis involves the identification of codes and categories representing themes that characterize the data being analyzed (O'Connor & Joffe, 2020). This human-performed, time-consuming task of coding and categorization can be performed with more accuracy and consistency, if it is done properly, using automated tools (Crowston et al., 2012). When large sample sizes are involved,

computer programs are able to perform many of the roles that would be prohibitive for researchers otherwise, due to the labor consuming nature of qualitative research (Crowston et al., 2012; Guetterman et al., 2018). The quantification of the frequency of words (or terms) does not necessarily transform the text into numbers, as was initially used in quantitative studies, but can rather be used to establish the importance of such terms (Vaismoradi & Snelgrove, 2019; Yu et al., 2011). Despite the initial preconception of computer-assisted text retrieval and analysis as purely quantitative tools, such processes are also valid methods for qualitative analysis, as the task of extracting and analyzing text content is indeed qualitative in nature, regardless whether such processes are performed entirely by a human, or with the assistance of a computer (Krippendorff, 2019; Yu et al., 2011).

Qualitative Research

Qualitative research seeks for a contextual explanation and interpretation of a social phenomenon or situation described by humans, for which two main data approaches are used, the qualitative content analysis (QCA) and the thematic analysis (TA) (Marvasti, 2019; Vaismoradi & Snelgrove, 2019). QCA and TA are used to examine human textual expressions by recurrent tasks of data analysis, coding and categorization (Vaismoradi & Snelgrove, 2019).

QCA deals with the systematic counting of text units and classification conducted using a methodological approach (Marvasti, 2019). In contrast, TA is concerned with the thick reflection of descriptive and interpretative nature, with specific focus on intentions, abstract themes and hidden agendas (Vaismoradi & Snelgrove, 2019). In fact, TA is interpretative in concept, relying on the "researcher's subjectivity and personal insight" (Vaismoradi & Snelgrove, 2019). QCA is often used as an initial step in data analysis, allowing the research to focus on the most frequent terms (Vaismoradi & Snelgrove, 2019).

Both approaches, QCA and TA, are powerful and can lead to insightful analyses. A disadvantage of QCA, when used in isolation, is the risk of the results being out of context

if no extra steps are taken. On the other hand, a disadvantage of TA, if not assisted by computers, is its limitation to smaller sample sizes due to the labor-intensive task of thick interpretation required (Vaismoradi & Snelgrove, 2019).

Text Mining

The term *text mining* was defined by multiple scholars as a process of extraction of useful data from documents through pattern identification, which comprises two components: *text retrieval* and *text analysis* (Crowston et al., 2012; Feldman & Sanger, 2009; Guetterman et al., 2018; Rose & Lennerholt, 2017; Sattikar & Kulkarni, 2012). Text retrieval, a process of extracting features from words, sentences, paragraphs, documents or even entire repositories, is widely used in quantitative studies (Rose & Lennerholt, 2017). Such features could be individual words in a specific language or corpus, or terms composed of one or multiple words that represent a single expression (Feldman & Sanger, 2009). When text retrieval and analysis are performed with computers, such processes significantly improve the quality of the inferences produced and reduce bias, all this due to the fact that software tools work in a consistent, methodological manner (Janasik et al., 2009; Yu et al., 2011).

Four-Stage Roadmap

Mixed methods designs typically contain one or more qualitative components and one or more quantitative components (Creswell & Plano Clark, 2006). Mixed-method designs with just two components are *sequential*, if the qualitative process is performed upon the findings of the quantitative components, or vise-versa. More complex designs may contain more than two components and a diverse combination of order between qualitative and quantitative processes. If the design is *concurrent*, quantitative and qualitative processes are performed independently, and the findings from each are combined (or merged) after both are complete (Creswell & Plano Clark, 2006).

These mixed methods designs assume that the components can be deterministic and categorized as either quantitative or qualitative. However, when some of the tasks, that are

traditionally performed by humans, are replaced by computer algorithms using statistical modeling (which usually is the case when text-mining is applied), this quantitative/qualitative distinction is no longer possible. When the components contain an intertwining combination of quantitative and qualitative parts, the design is considered *hybrid* by some scholars (Fahmi et al., 2018; Raich et al., 2014).

Research design for this study. The research design of this dissertation is an adapted version of the *Exploratory Design for an Instrument Development Model* published by Creswell and Plano Clark (2006), containing four stages (see Figure 4) comprising hybrid components, and based on the *IPOR Instrument Development Framework* (Nehoran & Nehoran, 2020). Therefore, I determined that the design of this dissertation is: *Mixed methods exploratory design for an instrument development.*

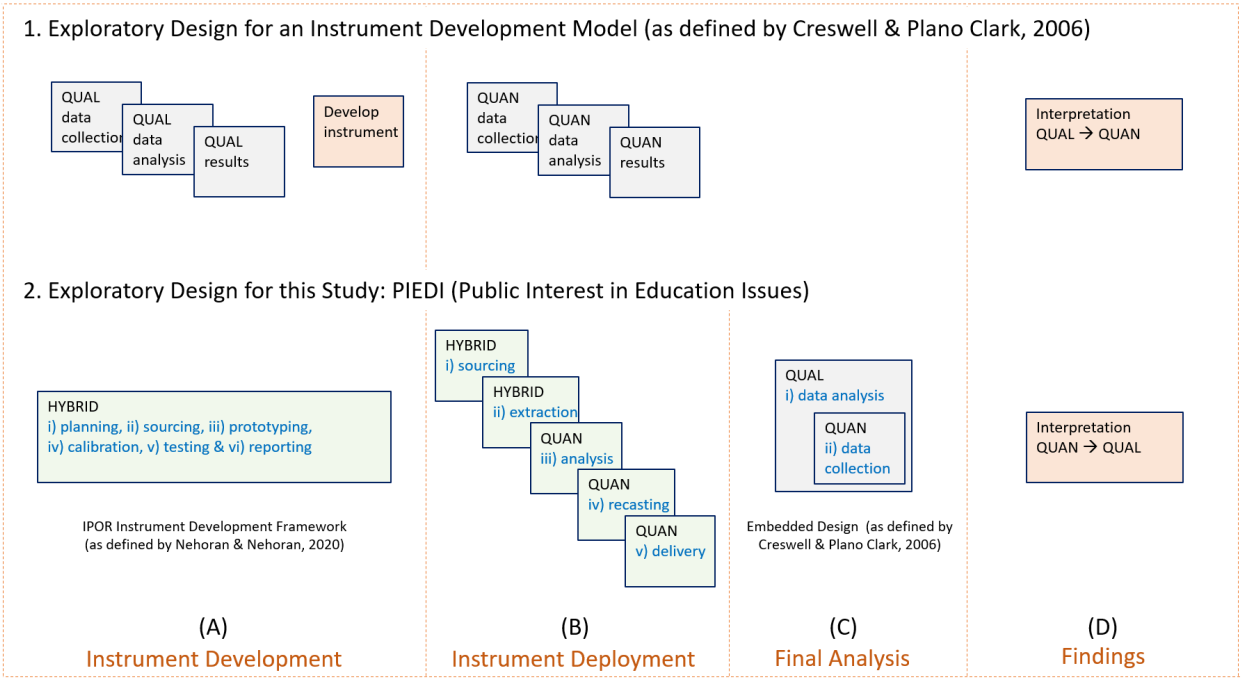


Figure 4. Dissertation research design & roadmap

Stage (A) - Instrument Development

The first stage (denoted "A") comprises the creation of the PIEDI instrument, which corresponds to the first stage of the *Exploratory Design for an Instrument Development Model* by Creswell and Plano Clark (2006). In the model by Creswell and Plano Clark (2006, p. 77), researchers develop a quantitative instrument based on qualitative findings to guide the development of items and scales for the survey (the instrument). Similarly, in this study, a quantitative instrument was built based on qualitative findings from an automated (computer performed) qualitative content analysis (QCA) to guide the development of the search queries (items) and related parameter adjustments (scales), both part of the newly developed software instrument.

To create time series sequences of the topics analyzed in this study, it is necessary to determine in advance the granularity of time period that will be used for the analysis. Using a high granularity scheme (e.g., weekly or monthly) would cause an insufficient amount of news articles published by single newspapers in each period. On the other hand, utilizing years as the period of choice would lead to a very small number of data points since 2015 (just six periods). For that reason, I selected quarters as the time period for this study.

The development also involves multiple cycles of instrument calibration, as well as reliability tests and validity assessment, here performed by comparing the results with a benchmark survey. The stage concludes after the software instrument created is able to: 1) collect the news articles meeting the search query that were published by the selected gatekeepers, 2) analyze the documents retrieved and extract the themes, and 3) deliver the expected findings (Nehoran & Nehoran, 2020).

Stage (B) - Instrument Deployment

Once the first stage is complete, the PIEDI instrument is ready to be used to collect the data needed for the analysis of fluctuations of interest in education issues in the US, through multiple processes of qualitative content analysis (QCA). After the data are

collected and analyzed, the PIEDI instrument produces a time series of the fluctuation of interest of issues in the US since 2015, which is the answer to the first question of this study.

Stage (C) - Final Analysis

The third stage ("C") corresponds to a typical simple *Embedded Design* as delineated by Creswell and Plano Clark (2006). To answer the rest of the questions of this study, it is necessary to select the documents pertaining to the highest fluctuations of public interest in education (a quantitative automated process performed by the instrument), and to conduct a human-performed thematic analysis on such documents.

In this stage a purposeful sample of documents related to the highest fluctuations of interest in education is collected. The final findings of this study include the aggregation and interpretation of the factors associated to the highest fluctuation of interest in education in the US, based on a thematic analysis (TA) conducted on the sample of documents collected.

Stage (D) - Findings

The fourth and last stage ("D") encompasses the interpretation of the findings and conclusions (typical chapters four and five of the dissertation), corresponding to a similar stage specified by Creswell and Plano Clark (2006).

Methodology

The utilization of aggregated-knowledge expert interviews as a collection method, used to understand the opinion of individuals through the perception of leaders, representatives or field professionals, is not uncommon among scholarly research. For example, teachers can be interviewed to understand what they think could help students learn math faster, instead of (or in addition to) interviewing individual students. Parents are often interviewed to learn about their children's feelings, instead of interviewing the children directly. A radiologist can be interviewed to understand their perception on how

comfortable the patients feel with a newly introduced medical instrument, instead of interviewing each patient separately. If the topic of research requires information about traumatic events (e.g. child abuse), interviewing counselors instead of the victims reduces the possible impact individual interviews could bring. Aggregated-knowledge expert interviews are a non-intrusive way of gathering the information, and are often used as methods of triangulation to increase the validity of the study of individual opinions.

Aggregated-knowledge experts (teachers, parents, political leaders, newspaper gatekeepers, etc.) acquire their knowledge via a constant interaction with multiple individuals. They are able to learn behaviors, preferences and sentiments through continuous observation over time. The more people these experts have access to, the more knowledge they accrue.

Newspaper Gatekeepers as Aggregated-Knowledge Experts

Newspaper gatekeepers are responsible for establishing the media agenda and making publishing decisions accordingly (Hodgkiss, 2017; Serban, 2015; Strömbäck et al., 2012; Vu, 2014). To maximize the economic gain of the published newspapers, gatekeepers consult a plethora of sources that provide them real-time readership information (Christin, 2018; Tandoc, 2014). Consumption dashboards, reports, online reader responses and blogs are just a few of the tools utilized by gatekeeper as a barometer of public opinion (Vu, 2014; Welbers et al., 2016). Experienced gatekeepers understand the opinions of their readers and act upon them by publishing material of interest to them, and thus representing their opinions (Vu, 2014).

The utilization of newspaper gatekeepers to understand the public interest in education issues is not new. In fact, in the 70's, Jacobson (1973) analyzed the news coverage as perceived by 301 media education writers to find the "ways of improving the coverage of education news". Later on, in the early 80's, Hynds (1981) interviewed 186 gatekeepers to explore the agenda-setting salience of education in the newspapers of that epoch. In other words, the researchers analyzed how much coverage education issues had

among all other issues covered by the newspapers then. Similarly, this study also analyzes the opinions of gatekeepers on education, but instead of interviewing a handful of them, it infers their opinions through the frequency of the terms utilized by them in tens of thousands of published news articles.

Design: Phenomenological Longitudinal Exploratory Mixed Methods

This study is categorized as a phenomenological exploratory mixed-method research. In this research, instead of using in-depth unstructured interviews, as it is typical in phenomenological studies, I conducted a thorough text analysis of the documents maintained by the gatekeepers to explore their perceptions on top-priority issues over the years. The findings of this study were reached through the utilization of the PIEDI instrument, built in this dissertation to collect and analyze the text of thousands of news articles related to issues of national priority in the US.

Description of the Participants

The data comprised all the gatekeepers working for one of the newspapers recognized for archival by the Library of Congress ("Library of Congress," 2020). At the time of retrieval, in January 2020, the list contained 414 newspapers, some of which shared web domains, and hence were considered a single entity in this study.

Since the gatekeepers working for a single newspaper cannot be separated, it is not possible to know the total number of individual gatekeepers that are represented in the sample of this study (some newspapers are large and have many gatekeepers, while others only have a handful). Therefore, this study considered groups of gatekeepers instead, each group representing the collective opinion of all gatekeepers working for a single newspaper.

After removing newspapers with no online access, and the ones written in a language other than English, the final selection for this study contained 258 groups of gatekeepers, each such group representing all the newspapers that share a web domain (typically, due to common ownership).

CHAPTER 4: FINDINGS

The findings of this dissertation first included the development of the popular Interest in Education Issues (PIEDI) software instrument able to collect tens of thousands of articles from the web, conduct a textual analysis and produce (after the instrument was successfully tested for reliability and validity) a set of reports that help answer the research questions of this study.

Once the instrument was ready, and could produce the necessary reports, the next sets of findings of this study included the full responses to the first two research questions and partial quantitative responses to the last two questions.

Lastly, a thematic analysis conducted over a set of articles (previously selected by PIEDI) complemented the quantitative responses to the last two questions with a detailed interpretation of the news covering the specific topics related to the largest fluctuations of interest in education.

(A) Instrument Development

Similar to a traditional survey instrument with an open-ended question, which collects the information from the respondents (in their own voice), Inferential public opinion instruments (IPOR-based) use a search query and a web search using that query to collect raw text documents from the groups of gatekeepers (Nehoran & Nehoran, 2020) in the voice of journalists. The instrument development of this study corresponds to the Exploratory Design for an Instrument Model by Creswell and Plano Clark (2006) (see Figure 4 on page 47).

As part of this dissertation, I created an IPOR-based instrument called PIEDI, which collected news articles related to national issues in the US from selected newspapers, analyzed the text in these articles, and delivered a time series of the fluctuations of public interest in education issues in the US in comparison to other issues of national importance. The steps of the software development process are delineated in this section.

The following sections will cover the five hybrid sections of PIEDI: i) planning, ii) sourcing, iii) prototyping, vi) calibration, v) testing and vi) reporting.

i) Planning

The planning phase of an IPOR instrument development involves the three main steps: a) establishing a *verifiable derived* research question and a corresponding *web query* to be used in document search, b) ensuring access to a document repository relevant to the verifiable research question and c) conducting a survey or finding an existing survey that can be used as the benchmark at the time of instrument validation (Nehoran & Nehoran, 2020).

Establishing a verifiable derived research question. The first research question sought an understanding of the fluctuations of public interest in education issues among the rest of the issues in the US since 2015. To answer the research question, a broader verifiable question was used, the most important problem (MIP) (which asks for all issues in the US - not necessarily education), to enable comparison with a benchmark survey used as a gold standard. The corresponding *web query* for this question is "issues in America," which resulted in the list of newspaper articles related to problems in the US, representing the media exposure.

Ensuring access to a document repository. I created a software tool that included a web crawler able to conduct a web search using the Microsoft Bing search engine and the web query specified above to retrieve relevant newspaper articles. The tool then extracted and refined the text to obtain the content of the news.

As part of the text refinement process, the software tool was programmed to remove from the collected news articles words, terms, sentences or entire paragraphs that are irrelevant to the objective of this research. Advertisement, email addresses, numbers and dates are all examples of irrelevant text that was removed from the documents to increase the accuracy of the results (see *Appendix A*). It is a common practice in text mining and natural language processing to remove numbers and terms that do not contribute to a

discriminant function that could divide documents into topics. For example, the term *january* does not help determine if the article was related to politics or sports.

The tool returned the refined text of all articles meeting the web query which were published by the selected newspapers within a specified time period. More specifically, the results included documents related to all issues in America, as monitored by the newspapers recognized for archival by the "Library of Congress" (2020).

Finding a benchmark survey. A number of US polling agencies with a long history of conducting surveys regarding the most important problem (MIP) were considered as a source for a benchmark survey for this study. After thoroughly evaluating the most renowned ones (Reuters, Associated Press, Gallup and Pew Research Center), I selected Reuters ("Reuters Poll Explorer," n.d.) as the agency of choice for this study. My decision was based on Reuters' uninterrupted record and frequency of the published survey data. Reuters polls are conducted (utilizing online interviews) by Ipsos and are typically called "Core Political Data" polls by "Reuters/Ipsos". Some examples of such reports are *Core Political Data Ipsos Poll Conducted for Reuters* (2016) and *Core Political Data Ipsos Poll Conducted for Thomson Reuters* (2020).

As specified in Reuters' Methodology page ("Reuters Poll Explorer Methodology," n.d.), Reuters has about 132,000 unique annual responses, and has been conducting weekly polls on public opinion (in a variety of forms) since 2012. The MIP question is aggregated regularly (every four weeks) covering topics of contemporary public interest. The exact text of the question is, "*In your opinion, what is the most important problem facing the US today? (Select from below or write in)*". The answers are selected from a predetermined list, but a space for a user-provided response is also available. The predetermined list, which did not change since 2012, contains the following options: *The Economy*, *Immigration*, *Energy issues*, *Crime*, *Unemployment/lack of jobs*, *Terrorism*, *Morality*, *The Environment*, *War/foreign conflicts*, *Healthcare*, *Education*, *I don't know*, *Other*. When a participant uses the open-ended section to provide a custom answer, the response is marked as *Other*.

I utilized the aggregated time series representation of issues in America published by Reuters ("Reuters Poll Explorer," n.d.) as the benchmark for this study. Each data point reported in Reuters' historical view is supported by a methodology page which explains the sample size (typically, a random sample of about $1,000$ participants), the demographic/socioeconomic composition of the respondents and the methods used to analyze the data. An example of such page is *Reuters/Ipsos/UVA Center for Politics State Poll: Tennessee* (2018).

ii) Sourcing

The data collection started with the selection of qualified news outlets (see *Appendix A* for details). All articles recognized for archival by the Library of Congress were considered for this study. After the qualification criteria was applied, a total of 258 newspapers were selected. See *Appendix B* for a complete list of articles selected.

The first step in the development of the PIEDI instrument was to create a software tool able to extract the content from the newspaper articles published in the year 2015 or later. The tool returned the refined text (cleaned from advertisements and auxiliary information) of all documents meeting the web query *"issues in America"*.

The web crawler collected $113,457$ newspaper articles based on the results of the web query in the Bing search engine. After the articles were qualified based on predetermined criteria utilized to filter out irrelevant ones (see *Appendix A* for details on the qualification criteria), the total number of articles remaining in the selected sample for this study was $51,262$.

iii) Prototyping

The prototyping phase of the instrument development included the completion of the software program, which was adapted to be able to select the most significant terms (based on press exposure) and calculate frequencies into two tables that are later used for analysis purposes.

Token extraction. The text in written documents is structured according to the syntax of its language and the terminology typical to its subject domain. For example, documents related to medicine are likely to include words and associations that are very different from the terms and associations found in documents related to immigration laws. Each subject domain has its own glossary of terms and association between them that fits the knowledge base and expectation of the target readership. The same word may have a completely different connotation when it is utilized in one domain than when it is used in another one. For example, the word *course*, when utilized in education articles is likely associated with other words related to a subject of instruction, in politics or social science contexts, it implies progression (e.g. *course of action*, or *course of history*), while in culinary literature it will be associated with a dish or a set of dishes served together. When two or more words are combined based on their part of speech, they form a *term*.

The text of the whole document can only be analyzed after a preprocessing task has been completed. Such task includes, among other steps, the *tokenization*, which extracts the tokens (or words) from the text, the removal of punctuation, and the transforming of the text into lowercase (Allahyari et al., 2017).

Terms. Lowercase tokens are grouped into terms (e.g., *state of the art*) based on their part of speech. It is up to the researcher to determine how terms are composed. For the development of the PIEDI instrument, I defined a *term* as a set of adjacent adjectives and nouns (including proper nouns). Other researchers willing to replicate this study may consider a different definition of a term. For example, a sentence including the text *The vice president of the United States of America*, leads to three terms: *vice president*, *united states* and *america*, while another researcher may decide to make it one single term containing all these tokens. Since my objective was to conduct a topic modeling analysis with the terms found, smaller terms increased the chances of higher frequencies of terms and the statistical significance of these frequencies for analysis. Regardless of the choice made, all terms will end up clustered together, adding to the load value of the same topic.

Basic unit of measurement. The basic unit of measurement for text mining projects is the frequency of a term in a document (*freq*), which calculates the number of times a term appears in a document.

If j is the number of the document, then d_j is the j^{th} document, then if the frequency of a term (e.g., *slavery*) in the document is defined as:

$$freq(slavery, d_j) = 4 \tag{1}$$

it means that the term *slavery* appeared four times in the document.

Similarly, if the frequency of a term (e.g. *brown university*) in the document is defined as:

$$freq(brown\ university, d_j) = 1 \tag{2}$$

it means that the term *brown university* appeared one time in the document.

Adjusted term frequency. Using the *freq* to measure prominence of terms is a good initial step, but very biased, since it gives undue importance to unusually long documents. For that reason, *freq* is seldom used as the basic unit of measurement for text mining projects. To address this issue, in this study, the adjusted term frequency (*atf*) was adopted, instead.

As shown in Equation 3, *atf* is obtained by dividing *freq* by the total frequency of all terms.

For document d_j, a term t_i and a total of k unique terms in the document d_i, the *atf* of the term is defined as:

$$atf(t_i, d_j) = \frac{freq(t_i, d_j)}{\sum_{j=1}^{k} freq(t_i, d_j)} \tag{3}$$

See Equation 4 and Equation 5 for examples of *atf* values. If the sum of frequencies of all terms in a document (e.g. d_j) was 71, then:

$$atf(university, d_j) = \frac{5}{71} = 0.0704 \tag{4}$$

$$atf(slavery, d_j) = \frac{4}{71} = 0.0563 \tag{5}$$

The sum of the *atf* of all terms in a single document always equals to 1.

***atf* table.** With a total of n documents containing k unique terms, a sparse table (containing mostly zeros) with n rows and k columns was created in which each cell was populated with the *atf* value of each document and term. Each row in the *atf* table sums up to 1 and each column sums up to a relative measure of salience for each unique term. The larger the relative measure, the more salient that term is among all other terms in the documents.

Significant terms. The number of unique terms k in this repository is very large (close to 1 million unique terms), and most of them appear just a handful of times. Since amount of computational power required for such a large table is very high, it is necessary to concentrate on the most salient terms, otherwise conventional computers cannot process the terms efficiently.

For this study, I have decided to concentrate on the $10,000$ most significant terms, a number herein denoted k_{sig}. Therefore, the columns are ordered by their sums, and only the top $10,000$ columns/terms are kept for further analysis (and the rest is discarded).

Sampling documents. Processing a very large number of documents with $10,000$ terms requires a sizeable amount of computational power, computer memory and time to execute the processes required. A repository of $51,262$ documents is considerably large, and some of the computationally intensive operations can take hours and even days to perform.

For this study, I have decided to select a random sample of $30,000$ articles out of the repository of $51,262$ documents. For reliability purposes, as explain later in this document, I repeated the process three times, to compare the results of the three samples utilized. I used the random sampling method (as typically utilized on clustering algorithms by data scientists), as opposed to the bootstrapping k-fold approach to ensure the size of each selection was large enough to meet the statistical requirements of this research.

Final dimensions of the atf table. The dimensions of the final *atf* table were $n = 30,000$ by $k_{sig} = 10,000$.

Quarterly *atf* table. The *quarterly atf table* contains m rows and $k_{sig} = 10,000$ columns, where m is the number quarters, and k_{sig} is the number of significant terms specified in the previous step. In this study I collected articles from 2015 to 2020, which led to a quarterly *atf* table with 22 rows ($m = 22$). Table 1 shows a sample portion of the quarterly *atf* table obtained in this study. The notation used for quarters is based on the four quarters within each year, with $Q1$ starting January 1. The quarter is abbreviated as $yyyyQq$ where $yyyy$ represents the year and q represents the quarter number (e.g., 2016Q1 covers the period starting on January 1, 2016, and ending on March 31, 2016).

Each cell of the *atf* table contains an aggregated measure of quarterly frequency, based on the documents published by each individual newspaper. This step is equivalent to interviewing the gatekeeper groups and asking them a question such as "What was the mean *atf* of the term *homelessness* in the first quarter of 2020 for all the articles published by your newspaper?"

For each newspaper (gatekeeper group) and unique term, I calculated the *means* of all articles published per quarter. Once the measure of means has been calculated, it was necessary to find a cutoff percentile where on one hand outliers were kept out of the calculation (which was achieved by using a percentile vs. mean), and on the other hand a statistical significant number of *mean atf* values were greater than zero. I found out that an aggregation using the 95^{th} percentile gave satisfactory results. In other words, for each

Table 1
Quarterly atf Table

Quarter	...	employ	program	russia	plan	questio	death	vote	candida	right	israel	change	virus	system	...
2015Q1	...	0.0028	0.0042	0.0010	0.0039	0.0034	0.0030	0.0021	0.0025	0.0041	0.0013	0.0037	0.0000	0.0031	...
2015Q2	...	0.0030	0.0060	0.0016	0.0048	0.0033	0.0021	0.0033	0.0037	0.0039	0.0008	0.0040	0.0001	0.0029	...
2015Q3	...	0.0025	0.0039	0.0008	0.0035	0.0041	0.0027	0.0031	0.0040	0.0034	0.0011	0.0028	0.0000	0.0031	...
2015Q4	...	0.0026	0.0059	0.0014	0.0040	0.0031	0.0019	0.0030	0.0047	0.0039	0.0008	0.0036	0.0000	0.0034	...
2016Q1	...	0.0021	0.0039	0.0012	0.0036	0.0030	0.0025	0.0046	0.0051	0.0060	0.0007	0.0049	0.0007	0.0043	...
2016Q2	...	0.0029	0.0041	0.0007	0.0040	0.0039	0.0030	0.0043	0.0056	0.0045	0.0018	0.0039	0.0009	0.0031	...
2016Q3	...	0.0037	0.0029	0.0027	0.0038	0.0043	0.0033	0.0028	0.0066	0.0042	0.0006	0.0029	0.0003	0.0027	...
2016Q4	...	0.0031	0.0032	0.0052	0.0050	0.0041	0.0035	0.0051	0.0064	0.0062	0.0015	0.0043	0.0000	0.0045	...
2017Q1	...	0.0037	0.0059	0.0036	0.0042	0.0066	0.0027	0.0037	0.0024	0.0044	0.0028	0.0039	0.0002	0.0029	...
2017Q2	...	0.0026	0.0033	0.0048	0.0041	0.0035	0.0026	0.0026	0.0026	0.0037	0.0008	0.0033	0.0000	0.0046	...
2017Q3	...	0.0026	0.0040	0.0052	0.0031	0.0052	0.0035	0.0040	0.0021	0.0049	0.0002	0.0032	0.0001	0.0035	...
2017Q4	...	0.0031	0.0037	0.0020	0.0035	0.0043	0.0032	0.0038	0.0034	0.0046	0.0026	0.0045	0.0003	0.0035	...
2018Q1	...	0.0029	0.0025	0.0016	0.0041	0.0030	0.0029	0.0033	0.0051	0.0048	0.0006	0.0046	0.0000	0.0031	...
2018Q2	...	0.0044	0.0037	0.0020	0.0046	0.0055	0.0025	0.0035	0.0046	0.0049	0.0025	0.0038	0.0005	0.0030	...
2018Q3	...	0.0028	0.0038	0.0033	0.0039	0.0044	0.0025	0.0031	0.0078	0.0045	0.0009	0.0038	0.0000	0.0034	...
2018Q4	...	0.0034	0.0035	0.0020	0.0059	0.0041	0.0020	0.0053	0.0040	0.0039	0.0007	0.0034	0.0000	0.0036	...
2019Q1	...	0.0037	0.0041	0.0009	0.0036	0.0034	0.0019	0.0039	0.0036	0.0041	0.0020	0.0039	0.0000	0.0037	...
2019Q2	...	0.0029	0.0054	0.0023	0.0044	0.0045	0.0047	0.0023	0.0041	0.0041	0.0014	0.0035	0.0002	0.0046	...
2019Q3	...	0.0025	0.0032	0.0016	0.0033	0.0034	0.0027	0.0036	0.0039	0.0065	0.0013	0.0034	0.0001	0.0033	...
2019Q4	...	0.0056	0.0039	0.0021	0.0057	0.0029	0.0018	0.0055	0.0047	0.0043	0.0007	0.0039	0.0000	0.0040	...
2020Q1	...	0.0044	0.0038	0.0012	0.0045	0.0050	0.0050	0.0036	0.0049	0.0038	0.0007	0.0030	0.0116	0.0039	...
2020Q2	...	0.0046	0.0026	0.0008	0.0035	0.0033	0.0058	0.0028	0.0022	0.0037	0.0005	0.0034	0.0086	0.0042	...
2020Q3	...	0.0030	0.0025	0.0019	0.0036	0.0027	0.0046	0.0029	0.0022	0.0042	0.0009	0.0028	0.0049	0.0027	...

quarter and unique term, I calculated the 95^{th} percentiles of the means of all newspapers. The *Quarterly atf table* was utilized later to scale and calibrate the instrument.

iv) Calibration

The calibration process of an IPOR instrument resembles a typical qualitative thematic analysis in concept, but the process used to obtain the codes, themes and interpretations for the instrument contain a combination of automated tools and human interaction.

The objective of the calibration process of the instrument is not to write new software, but to adjust numerical values that help correlate the results obtained from the news articles with the results reported by the benchmark survey. The calibration results in a function, created with human interaction, which is not necessarily linear.

The thematic analysis performed for the calibration, as typical to any IPOR-based instrument (Nehoran & Nehoran, 2020), is conducted with a) an initial quantitative *topic discovery* process using machine learning algorithms and natural language processing,

called *topic modeling*, b) a task of *human-assisted theme discovery*, a qualitative analysis process that follows the legacy process of a typical phenomenological research, and c) a *technical adequacy test*, a human-assisted formula creation that attempts to match the results from the benchmark survey with the results obtained from the news articles.

Topic discovery. In machine learning and natural language processing, *topic modeling* is a modeling technique used to uncover "latent topics" that appear in a collection of documents given the semantic structure of written text (Li & McCallum, 2006; Vayansky & Kumar, 2020). Latent Dirichlet allocation (LDA) is one of the most popular topic modeling algorithms, but it is not able to use correlations between the topics, which makes the technique difficult to use when some topics are correlated (Li & McCallum, 2006). To overcome the limitations of LDA, Li and McCallum (2006) introduced a brand-new technique called the *Pachinko allocation model (PAM)*, a method that gained popularity among the research community.

As with any other classification algorithms (e.g., factor analysis), topic modeling requires the determination of the number of topics desired k_{tm}. Multiple rounds of trying different k_{tm} values and assessing the results obtained may be necessary. At the end of the process, *topic loading matrix* of dimensions k_{sig} by k_{tm} is obtained, with k_{sig} representing the number of significant terms $(10,000)$ and k_{tm} representing the number of topics obtained. The *topic loading matrix* resembles the typical loading matrix of any exploratory factor analysis.

Human-assisted theme discovery. When applying topic modeling, the most significant terms of each topic (the terms with the highest loads) can give a hint about the theme, but cannot describe it fully.

While the topic modeling algorithm is able to group terms that appear in similar proportions in documents, the computer often does not have the capability of assigning a name to each topic (and hence, the topics are called *Topic 1*, *Topic 2*, and so on). It is easier for humans (by reading the terms with the highest loads for each topic) to make

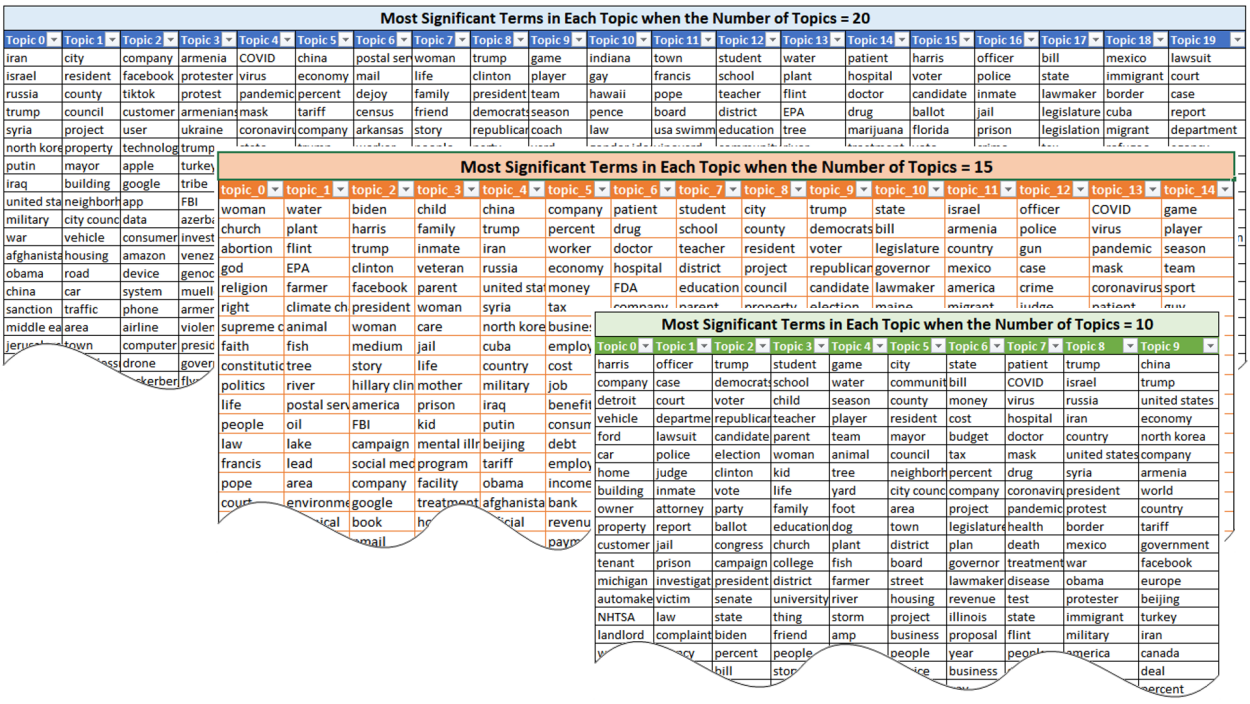
inferences and assign a label to each topic discovered. For example if a topic includes the term *production* with a weight of 0.27 , *GDP* with a weight of 0.19, and lower weights for the rest of the significant terms, it is fair to assume, by reading these terms, that this topic is related to the state of the economy and not to film production.

Technical adequacy test. The objective of a quantitative instrument is to obtain numerical values that reflect the data analyzed correctly. IPOR-instruments may be compared with a benchmark survey to ensure that their results correspond to the survey and hence can represent the public opinion (Nehoran & Nehoran, 2020).

The dot multiplication of the *quarterly atf table* by the *topic loading table* results in a set score coefficients that can be further adjusted by a function that matches the topic with that of the benchmark survey. To be able to perform that match, it is necessary that the topics obtained resemble the ones covered by the benchmark survey obtained.

Since k_{tm} (the number of topics that PIEDI builds) needs to be established upfront, I experimented with $k_{tm} = 10$, $k_{tm} = 15$ and $k_{tm} = 20$ with the objective of finding the topics that are as close as possible to the Reuters benchmark survey ("Reuters Poll Explorer," n.d.), which contains the following categories: *The Economy*, *Immigration*, *Energy Issues*, *Crime*, *Unemployment*, *Terrorism*, *Morality*, *The Environment*, *War/Foreign Conflicts*. *Healthcare*, *Education*, *Don't Know* and *Other*. Figure 5 shows the most significant terms (with the highest load) included in each topic obtained for $k_{tm} = 10$, $k_{tm} = 15$ and $k_{tm} = 20$.

The assignment of a label to each topic involves human interpretation of the most significant terms it contains. While the label selected is subjective, and different researchers may decide to name the topics differently, the results of PIEDI's analysis are not dependent on the label assigned, since the research questions of this study are answered based on the content of the news articles associated to each topic and not on its label.



Most Significant Terms in Each Topic when the Number of Topics = 20

Topic 0	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6	Topic 7	Topic 8	Topic 9	Topic 10	Topic 11	Topic 12	Topic 13	Topic 14	Topic 15	Topic 16	Topic 17	Topic 18	Topic 19
iran	city	company	armenia	COVID	china	postal serv	woman	trump	game	indiana	town	student	water	patient	harris	officer	bill	mexico	lawsuit
israel	resident	facebook	protester	virus	economy	mail	life	clinton	player	gay	francis	school	plant	hospital	voter	police	state	immigrant	court
russia	county	tiktok	protest	pandemic	percent	dejoy	family	president	team	hawaii	pope	teacher	flint	doctor	candidate	inmate	lawmaker	border	case
trump	council	customer	armenians	mask	tariff	census	friend	democrats	season	pence	board	district	EPA	drug	ballot	jail	legislature	cuba	report
syria	project	user	ukraine	coronaviru	company	arkansas	story	republican	coach	law	usa swimm	education	tree	marijuana	florida	prison	legislation	migrant	department
north kore	property	technolog	trump																
putin	mayor	apple	turkey																
iraq	building	google	tribe																
united sta	neighborh	app	FBI																
military	city counc	data	azerba																
war	vehicle	consumer	invest																
afghanista	housing	amazon	venez																
obama	road	device	genoc																
china	car	system	muell																
sanction	traffic	phone	armer																
middle ea	area	airline	violer																
jerusal	town	computer	presic																
		drone	gover																

Most Significant Terms in Each Topic when the Number of Topics = 15

topic_0	topic_1	topic_2	topic_3	topic_4	topic_5	topic_6	topic_7	topic_8	topic_9	topic_10	topic_11	topic_12	topic_13	topic_14
woman	water	biden	child	china	company	patient	student	city	trump	state	israel	officer	COVID	game
church	plant	harris	family	trump	percent	drug	school	county	democrats	bill	armenia	police	virus	player
abortion	flint	trump	inmate	iran	worker	doctor	teacher	resident	voter	legislature	country	gun	pandemic	season
god	EPA	clinton	veteran	russia	economy	hospital	district	project	republican	governor	mexico	case	mask	team
religion	farmer	facebook	parent	united sta	money	FDA	education	council	candidate	lawmaker	america	crime	coronavirus	sport
right	climate ch	president	woman	syria	tax									
supreme c	animal	woman	care	north kore	busines									
faith	fish	medium	jail	cuba	employ									
constitutio	tree	story	life	country	cost									
politics	river	hillary clin	mother	military	job									
life	postal serv	america	prison	iraq	benefit									
people	oil	FBI	kid	putin	consum									
law	lake	campaign	mental illr	beijing	debt									
francis	lead	social med	program	tariff	employ									
pope	area	company	facility	obama	income									
court	environme	google	treatment	afghanista	bank									
		book			revenu									
		mail			payr									

Most Significant Terms in Each Topic when the Number of Topics = 10

Topic 0	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6	Topic 7	Topic 8	Topic 9
harris	officer	trump	student	game	city	state	patient	trump	china
company	case	democrats	school	water	communit	bill	COVID	israel	trump
detroit	court	voter	child	season	county	money	virus	russia	united states
vehicle	departme	republican	teacher	player	resident	cost	hospital	iran	economy
ford	lawsuit	candidate	parent	team	mayor	budget	doctor	country	north korea
car	police	election	woman	animal	council	tax	mask	united states	company
home	judge	clinton	kid	tree	neighborh	percent	drug	syria	armenia
building	inmate	vote	life	yard	city counc	company	coronaviru	president	world
owner	attorney	party	family	foot	area	project	pandemic	protest	country
property	report	ballot	education	dog	town	legislature	health	border	tariff
customer	jail	congress	church	plant	district	plan	death	mexico	government
tenant	prison	campaign	college	fish	board	governor	treatment	war	facebook
michigan	investigat	president	district	farmer	street	lawmaker	disease	obama	europe
automake	victim	senate	university	river	housing	revenue	test	protester	beijing
NHTSA	law	state	thing	storm	project	illinois	state	immigrant	turkey
landlord	complaint	biden	friend	amp	business	proposal	flint	military	iran
	cy	percent	people		people	year	peopl	merica	canada
		bill	stor		ce	business			deal
									percent

Figure 5. Most significant terms of each topic

The labels obtained from PIEDI with $k_{tm} = 10$ were: *Housing/Manufacturing*, *Crime*, *Politics*, *Education*, *The Environment/Sports*, *Local Issues*, *The Economy*, *Healthcare*, *International Affairs* and *International Trade*.

The order of the topics obtained from the topic modeling algorithm is random (for example, *Education* is not more popular than *The Economy*, despite appearing earlier in the topic list). Only after the multiplication of the *Quarterly atf Table* by the *Topic Loading Table*, the topics can be ranked (quarter by quarter).

When experimented with $k_{tm} = 15$, the themes obtained seemed to be: *Religion*, *The Environment*, *Politics*, *Wellbeing*, *International Affairs*, *The Economy*, *Healthcare*, *Education*, *Local Issues*, *Elections*, *Legislation*, *Immigration*, *Crime*, *Public Health* and *Sports*.

Lastly, when I tried with $k_{tm} = 20$, the topics appeared to be: *International Affairs*, *Local Issues*, *Science & Technology*, *Russia*, *Public Health*, *International Trade*, *U.S. Mail*,

Morality, *Politics*, *Sports*, *Gender Identity*, *Other*, *Education*, *The Environment*, *Healthcare*, *Elections*, *Crime*, *Economy*, *Immigration* and *Legislation*

Selecting the right k_{tm}**.** As typical to any unsupervised clustering algorithm, I needed to select the most optimal number of clusters (in this case *topics*) based on criteria that met the objectives of this study. Since the intended result was to obtain a list of topics as similar as possible to the gold standard benchmark from Reuters/Ipsos ("Reuters Poll Explorer," n.d.), I compared the similarities that different k_{tm} produced. To match the topics produced by PIEDI and the ones reported by Reuters/Ipsos, a human interpretation of the topic names was required.

The Sankey diagram shown in Figure 6 shows the result of the human-performed comparison analysis I conducted when k_{tm} was 10, 15 and 20. The green bands show a direct match between PIEDI's results and Reuters/Ipsos' reports. As can be observed, with $k_{tm} = 10$, there were three direct matches: *Crime*, *Healthcare* and *Education*. With $k_{tm} = 15$, there were also three direct (but slightly different) matches: *Crime*, *Immigration*, and *Education*. Lastly, with $k_{tm} = 20$, there were four direct matches: *Crime*, *Immigration*, *Education* and *Morality*. The yellow bands represent partial matches between the topics. For example, with $k_{tm} = 10$, The *Economy* in PIEDI was matched to both *The Economy* and *Unemployment* in Reuters/Ipsos reports. Similarly, a partial match showed in $k_{tm} = 20$ with *International Affairs*, *Russia* and *International Trade* all had a partial match to a single Reuters/Ipsos' topic called *War/Foreign Conflicts*, while PIEDI's topic *International Affairs* also matched the *Terrorism* category of Reuters/Ipsos. As can be seen in Figure 6 and as expected, the larger the value of k_{tm} is, the more the number of PIEDI's topics is that partially match Reuters/Ipsos' *Other* category. The grey bands (at the bottom) represent no match.

In my opinion, Reuters/Ipsos question about the problems facing the US implies national issues, and people are not likely not consider local topics (such as *municipality*, *county*, *city council* or *mayor*) when they answer such question. In contrast, I found that

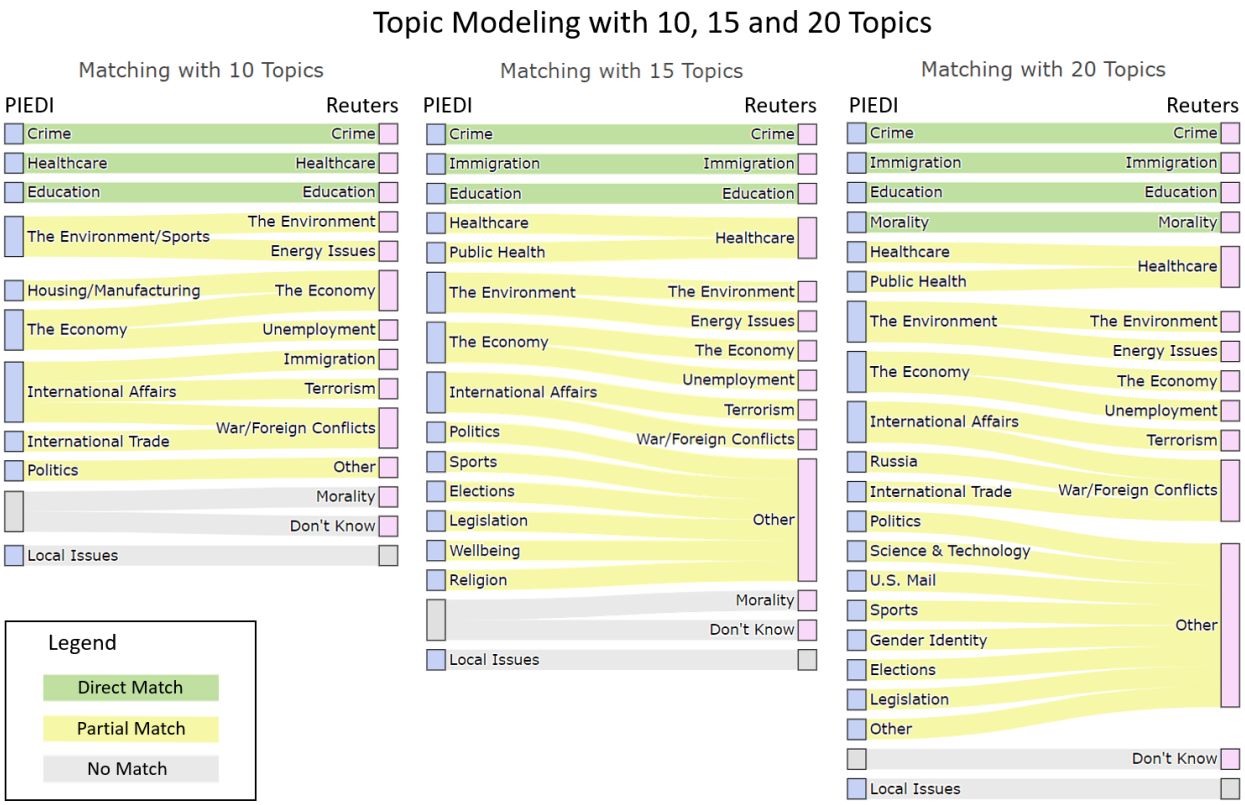


Figure 6. Comparison of topic modeling results for different k_{tm} values

local issues commonly appear in the news related to problems in America. I have therefore assumed that all problems related to *Local Issues* discovered by PIEDI are out of the scope of Reuters/Ipsos' poll and hence marked as *no match* in the comparison diagram.

There is no equivalent to Reuters/Ipsos' *Don't Know* answer in PIEDI. For that reason, it was marked as *no match*.

By analyzing the alternatives in Figure 6, I concluded that $k_{tm} = 20$ was the optimal value for this study because it produced the largest number of direct matches between PIEDI and Reuters/Ipsos, as well as a larger variety of topics for the *Other* category of Reuters/Ipsos, as expected from a poll that offers only 11 categories and a field for a user provided answer.

Education over time. The next step was to evaluate the fluctuations of interest in *education* since 2015, for which it was necessary to produce a temporal structure that showed education measures over time.

Education time series. The dot multiplication between each row in the *Quarterly atf matrix* and the *education* column of the *Topic Loading Table* results in a quarterly value of a time series called *Education Time Series*, with 22 data points, one for each quarter (see Figure 7).

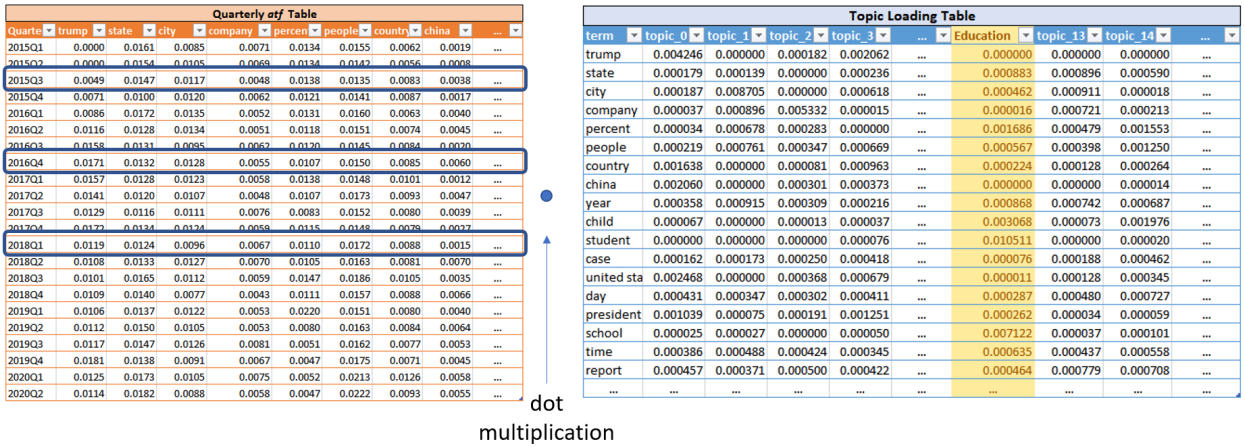


Figure 7. How Education Time Series is created (example)

Education relative time series. The absolute value of press coverage of education issues might be a valid measure, but may not be useful when the intention is to understand the percentage of interest education has among all issues of national priority, and its change over time. Additionally, this is important when comparing to the Reuters/Ipsos poll in which they report a relative score for *Education* against the rest of the national issues. For example, a value of 5.21 in the Reuters/Ipsos poll represents that 5.21% of the respondents mentioned "*Education*". Therefore, *relative interest* is implied in any mention of interest in issues throughout this study.

To obtain a relative scale of measurement, I have created the *Education Relative Time Series* which contains the coverage of education in the press divided by the sum of the coverage of all of the topics analyzed each period. The *Education Relative Time Series* represents a relative score against all other issues of national interest.

The calculation of the *Education Relative Time Series* (as shown in Figure 8) starts with the dot multiplication of the *Quarterly atf Table* by the *Topic Loading Table* (the entire table, not just education). This dot multiplication produces 20 time series sequences, one for each topic (including *Education*). The relative value of education in each quarter is obtained by dividing the *Education* value by the sum of all values in that quarter. This results in a time series for *Education*, that shows the relative value of interest in education as a percentage by quarter, when considering all other topics covered in the news.

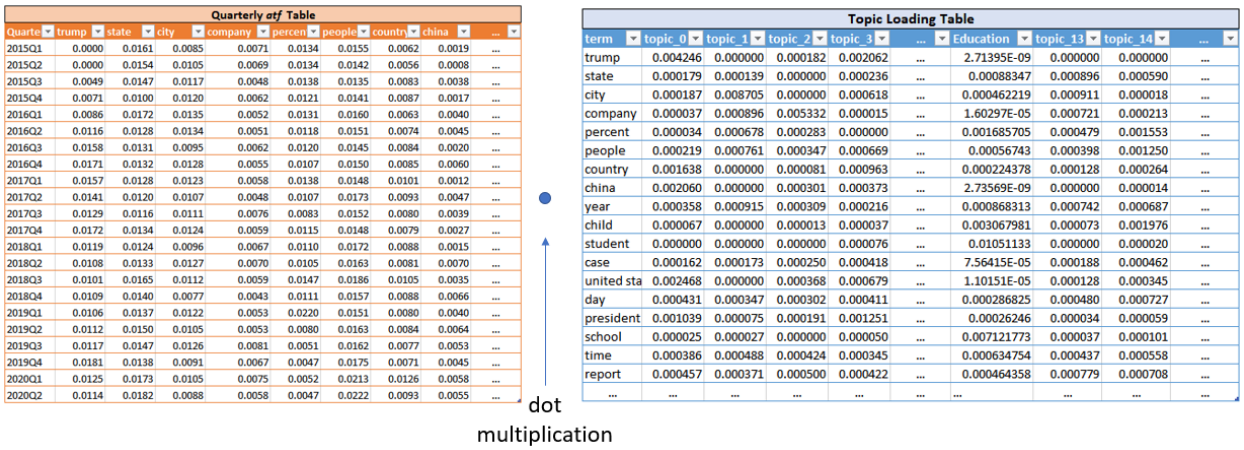


Figure 8. How Education Relative Time Series is created (example)

All topic difference table. A table containing the difference in quarterly values for all the 20 topics and 22 quarters (called *All Topic Difference Table*) was programmed in IPOR to produce the information necessary to answer the last three research questions. The first row of the All Topic Difference Table is empty, while the rest of the rows contain the difference between the quarter representing the row and the previous one. A large value

for a particular topic in a specific quarter (e.g. Education in Q2 2015) corresponds to a surge in interest. Similarly, a significant negative value indicates a plunge in that quarter.

v) Testing

Reliability. The *internal consistency* type of reliability deals with the measurement of the error that may be caused by the instrument utilized, and it is mostly concerned with the consistency of results when the tool is utilized repeatedly (Bannigan & Watson, 2009). The smaller the variation produced in repeated measurements, the larger the internal consistency reliability score of the instrument (Bannigan & Watson, 2009). As can be seen in Figure 9 and Figure 10, in this study I conducted an internal reliability test by running repeated executions of the instrument over different samples and then correlating the results.

Figure 9. Comparison between samples

The PIEDI tool was programmed to automatically select three equally sized random samples, produce their respective *Quarterly atf tables*, multiply these tables by the education column of the *Topic loading matrix*, and produce an education time series for each of such samples. The instrument then produces a plot for a visual comparison. More specifically, with sample sizes of $30,000$ documents, the tool was programmed to produce three different *education time series* one for each of the samples. The resulting time series are shown in the Figure 9.

While it can be visually observed in Figure 9 that the three samples follow a common pattern of fluctuation over time, a more deterministic statistical approach was also utilized to determine the correlation between the three executions. By comparing the quarterly resulting values from each of the samples, the tool was programmed to create the correlation matrix between the resulting time series lines (see Figure 10).

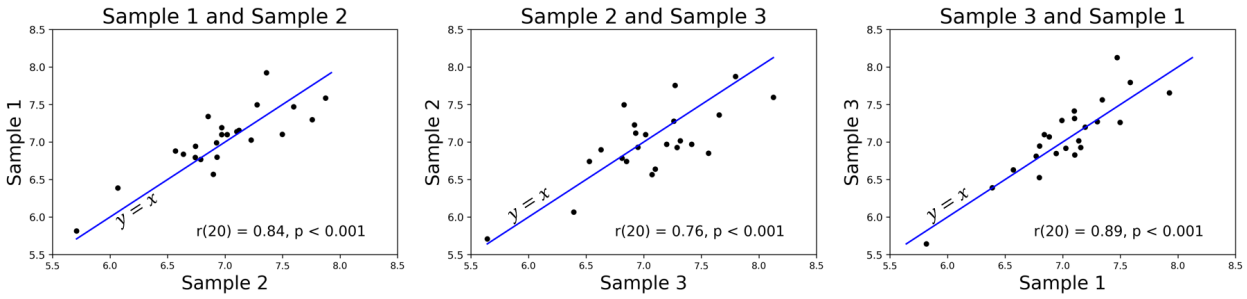


Figure 10. Correlation of samples

With a significance value of $p < 0.001$, the three time series sequences analyzed shown a statistically significance correlation between each pair of samples. As can be observed in the Figure 10, there is a Pearson correlation of 0.84 between the first two samples, 0.89 between the third sample and the first one, and a correlation of 0.76 between last two samples. These highly correlated scores between the samples show a high level of reliability of the PIEDI instrument. The tool was set to report, in addition to the two plots, the *mean reliability score*, which is calculated as the mean between all pair-wise

correlations. In this study the *mean reliability score* was 0.83, which is considered very good.

Validity. Concurrent validity is a test that evaluates the extent to which an instrument under development correlates with a gold standard, seeking high correlation scores between the two measurements (Bannigan & Watson, 2009). In the PIEDI instrument the intention is to find a statistically significant correlation with the reported results of a benchmark survey, and in this case the Reuters/Ipsos survey is used as the gold standard.

The *Education Relative Time Series* produced by PIEDI was based on the prevalence of education topics among all other topics covered in the news. In contrast the data published by Reuters ("Reuters Poll Explorer," n.d.) was collected by recurrent internet surveys (conducted by Ipsos) and showed the importance given by the survey participants to education (among the all other issues of national interest), when they are asked the MIP question. The focus of this validation process is to evaluate the level of alignment between these two time series lines, despite the disparate sources each of them utilizes for data collection.

Reuters/Ipsos surveys are conducted online and published every four weeks. They include the following MIP question: "*In your opinion, what is the most important problem facing the US today?* (Select from below or write in)". Each poll is administered over a representative sample of about $1,000$ participants. Their data was reported with exact four-week regularity from 2012 to early 2019, a year during which the periodicity revealed some irregularity with larger or smaller periods of time. For that reason, the validity of this research was conducted by comparing the data from 2015 (which is the starting point of the collection of newspaper articles for this study) to the end of 2018. The data from Reuters/Ipsos was resampled to quarters by taking the median value of the three four-week values falling within each quarter to allow the comparison with PIEDI, yielding a total of 16 quarters of common information.

PIEDI's *Education Relative Time Series* shows on top and the one reported by Reuters are plotted in the bottom of Figure 11. Although the exact measures by PIEDI could have been calibrated to match the range of Reuters/Ipsos's results, such process was not necessary here because this study concentrated on fluctuations, also known as slopes, and not on the quarterly values.

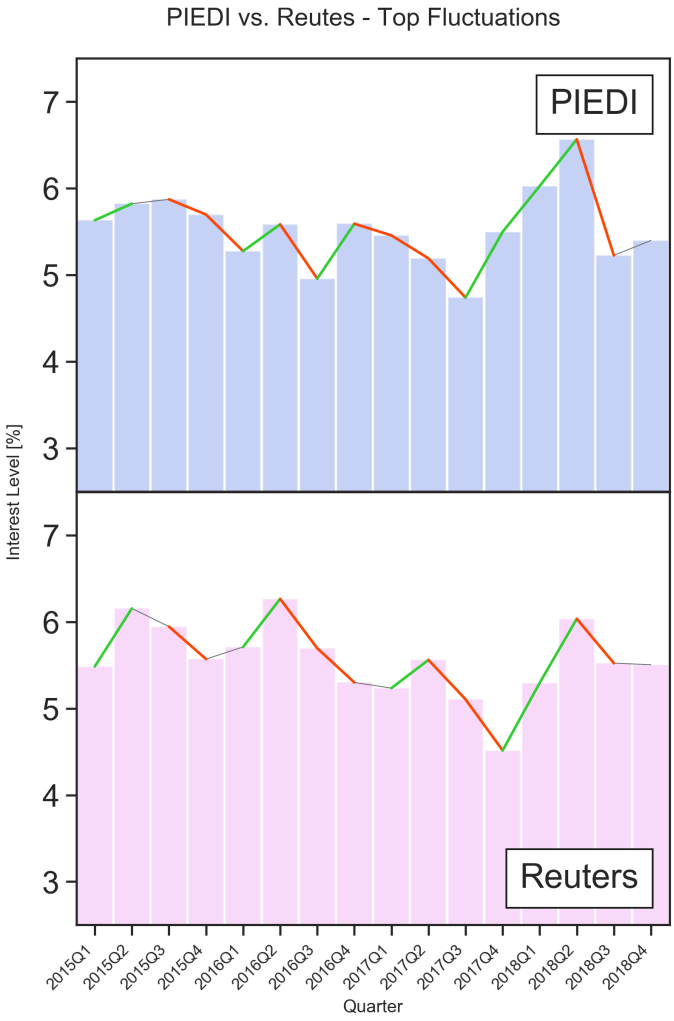


Figure 11. PIEDI vs. Reuters

Figure 11 enables a visual evaluation of the fluctuations of interest in education resulting from each method. The main fluctuations are highlighted in color (rises marked

as green and declines marked as red). As can be seen, there is an easily observable visual resemblance between the two series. A deeper visual analysis of the plot, comparing the two shapes, reveals that there are a few single-quarter shifts between the two series fluctuations. The literature review of this dissertation explains the mutual influence between the media and public opinion. When a topic gains prominence in the news, the public reacts and becomes more interested in that topic and vise-versa. For example, this could potentially be the case in Q3 2017, when the media exposure of education experiences a sudden growth, which was not observed in public opinion polls until the following quarter.

Dynamic Time Warping (DTW). Visual evaluation, as a method used to analyze a general alignment of time series lines, can be used as a first step, but a more rigorous and statistically-based method is able to provide a more quantitative approach to such alignment observed visually. The dynamic time warping (DTW) algorithm attempts to find patterns to match (and sometimes subsequently cluster) temporal sequences by comparing point-to-point distances (Jiang et al., 2020; Senin, 2008; Yuan et al., 2019). Traditional DTW algorithms work well to compare distances between two series, but are not sufficient when the objective of the task is to compare slopes (and consequently shapes), which is the objective in this study.

For this study, since I needed to correlate the alignment of shapes, I utilized an extended version of DTW called *locally slope-based DTW* (Yuan et al., 2019), which compares between locally similar slopes for each data point. The first step of *locally slope-based DTW* is to match common areas of concavity and convexity (see *Step 1* in Figure 12). It then matches the points to align the common slopes (see *Step 2* in Figure 12) to ultimately create a list of matching points using interpolation between adjacent extremes, when necessary (see *Step 3* in Figure 12).

The number of matching data points obtained by the *locally slope-based DTW* may be larger than the original list due to the additional interpolated points added. The correlation between the points in the *list of matching points* produced by the *locally*

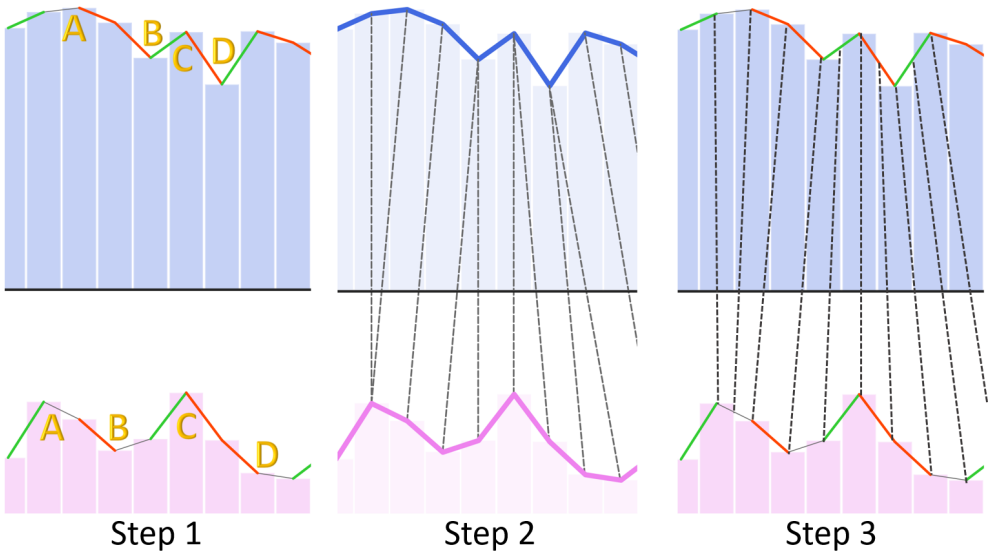


Figure 12. Locally Slope-Based Dynamic Time Warping 3-step process

slope-based DTW gives an indication of shape similarity between the two time series sequences. The higher the correlation is, the closer in shape the two time series sequences are.

As can be observed in the Figure 13, when *locally slope-based DTW* was applied to compare the 16 data points between PIEDI and Reuters/Ipsos, a clear matching pattern between the sequences was found.

Figure 13 shows a clear alignment of shapes (and hence slopes) between the two sequences and provides a *list of matching points* containing 19 data points.

Change from previous quarter. The *Change from Previous Quarter* is calculated by subtracting the percentage of the topic had in the prior quarter from the percentage in the quarter being analyzed. For example, if a topic was 0.48% in the prior quarter and it is 0.70% in the quarter being analyzed, the *Change from Previous Quarter* would be 0.22. A negative change indicates a downward slope, while a positive change denotes an upward slope.

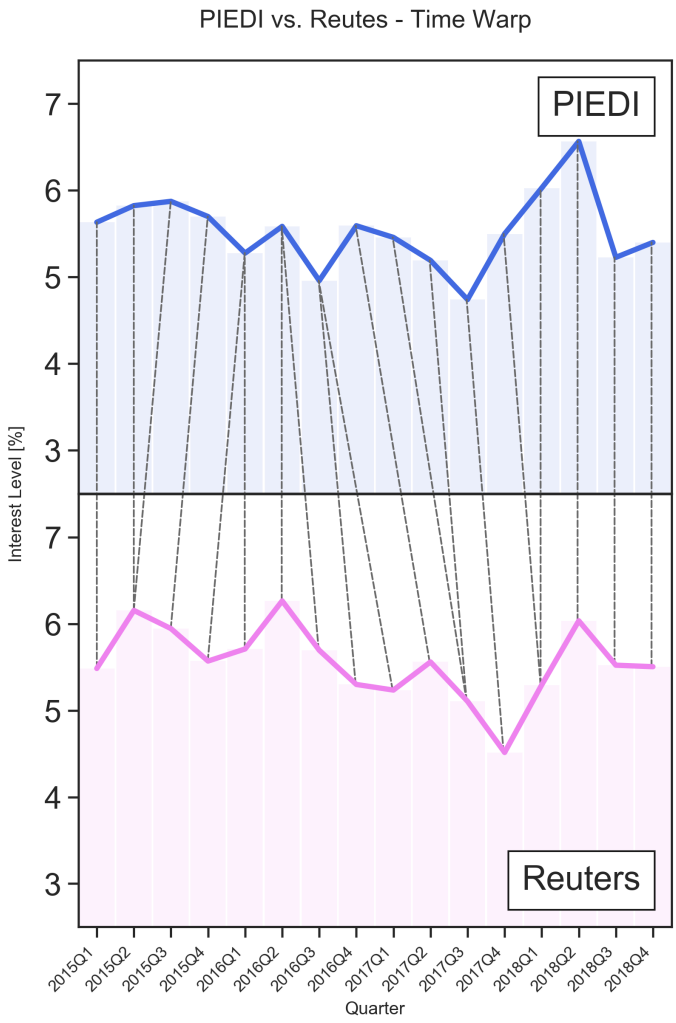


Figure 13. PIEDI vs. Reuters: Locally Slope-Based Dynamic Time Warping

Pearson correlation. A Pearson correlation was run using the *Change from Previous Quarter* values, applied to the *list of matching points* produced by the *locally slope-based DTW* (see Figure 14).

With a $p < 0.001$ and a correlation of $r = 0.79$, I was able to assert that the fluctuations of interest in education discovered by PIEDI were statistically similar to the ones reported by the Reuters/Ipsos polls, and therefore conclude that PIEDI was a valid instrument that could be used as a barometer to monitor fluctuations in public interest in education over time.

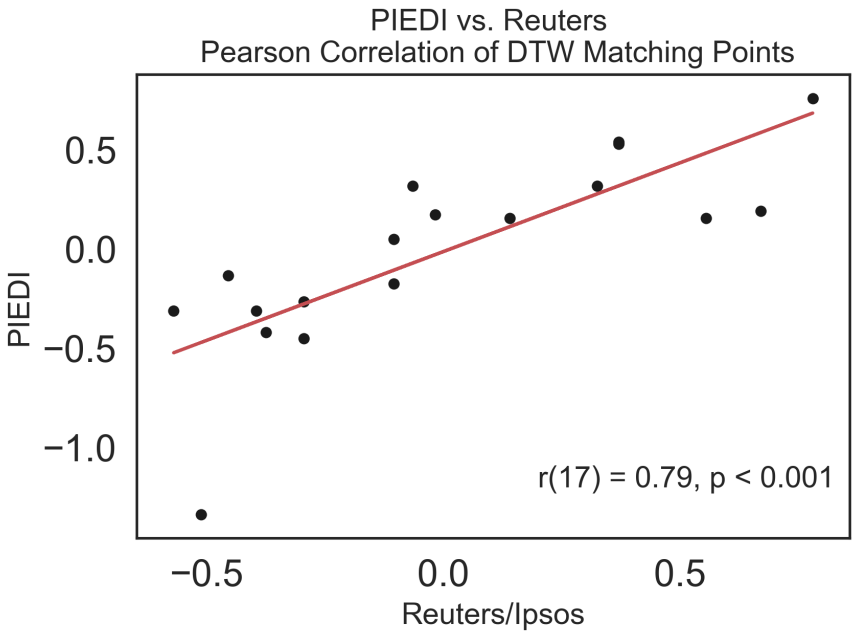


Figure 14. PIEDI vs. Reuters: Pearson correlation after Locally Slope-Based DTW

vi) Reporting

The PIEDI software instrument was evaluated for reliability to ensure consistency of results when applied to different samples, and for validity against Reuters/Ipsos polls to corroborate that the tool can be used to measure the public interest in education. The reliable and valid tool was then programmed to produce the reports necessary to answer the four research questions.

Document annotation: quarterly cohorts of relevant annotated articles. A human-performed thematic analysis of the articles related to each topic during a particular quarter helps identify the issues the topic represents. To assist in the selection of articles to read and analyze qualitatively, I programmed PIEDI to create cohorts of annotated documents with highlighted text. After the human-performed thematic analysis was conducted, I was able to complete the answers to the third and fourth questions of this study.

Summary of Instrument Development

The IPOR-based software instrument developed in this study, called Popular Interest in Education Issues (PIEDI), was utilized to analyze the text of tens of thousands of news articles related to problems in the US. The tool was programmed to discover the 20 latent topics that represent the public interest in America, and to later discover the 13 subtopics of education issues.

The instrument produced the *Education Relative Time Series* since 2015, which helped answer the first research question and a list of quarters with the most marked fluctuations of public interest in education to help answer the second research question.

In addition, to help answer the third and fourth question, PIEDI was programmed to produce a list of national topics that surged during quarters of education decline, and a list of education subtopics that surged during quarters in which the interest in education surged. To ease with human-performed thematic analysis that follows, PIEDI was programmed to produce cohorts of news articles related to each of the topics and subtopics published in the quarters in which the largest fluctuations of interest in *education* occurred. Only after the human-performed thematic analysis was completed, it was possible to fully answer the last two questions of this study.

(B) Instrument Deployment

The validated PIEDI software instrument created in this study was able to perform many of the labor-intensive tasks of data retrieval and analysis. The instrument was capable of retrieving the documents from the internet, performing an in-depth analysis, conducting reliability and validity tests, and producing a time series of fluctuations of public interest in education since 2015. It also produced sets of annotated documents relevant to the most prominent issues in each quarter of education interest rise or decline.

The deployment of PIEDI contains five phases: a) sourcing, b) extraction, c) analysis, d) recasting and e) delivery (see Figure 4 in page 47), most of which are executed

without human interaction. This section includes a detailed description of the delivery phase, which contains the results of this study.

i) Hybrid Sourcing

The PIEDI software instrument was executed with a period of time from 2015 to August 2020. The instrument automatically collected the news articles published by gatekeepers working at the selected newspapers. The instrument completed the hybrid sourcing and extraction components of the exploratory design for this study (see Figure 4, Dissertation research design & roadmap). No human interaction was required for this phase.

ii) Hybrid Extraction

The PIEDI refined the documents read from the internet (by removing advertisements, links, email addresses, and other irrelevant data). It also extracted the terms that would be later used to conduct the qualitative content analysis (QCA). No human interaction was required for this phase.

iii) Quantitative Analysis

The qualitative analysis task, which required the utilization of frequencies to uncover prevalent terms, and coding to group the terms into topics, was automatically performed by the PIEDI instrument, without human interaction.

iv) Quantitative Recasting

Recasting is the phase in which the numerical values are analyzed, including correlations, mathematical transformations and time series analyses. Neither correlations, nor numerical transformations were needed for this study. The recasting task created a time series analysis, all automatically performed by the PIEDI, without human interaction.

v) Quantitative Delivery: Quantitative Answers to Research Questions

The execution of PIEDI produced a set of results, including plots, lists and document cohorts, that helped answer the four questions of this study.

1. First research question: ***How has the interest in education, among all the national priority issues, changed in the US from 2015 to 2020?*** The execution of PIEDI produced a time series plot emphasizing the largest fluctuations of interest in education since 2015 (see Figure 15). The surges in interest were marked in green, and the plunges in red.

A visual representation of the time series helped evaluate the fluctuations of interest over time. The bold red lines show the steepest declines in relative interest in education, while the bold green lines show the steepest rises.

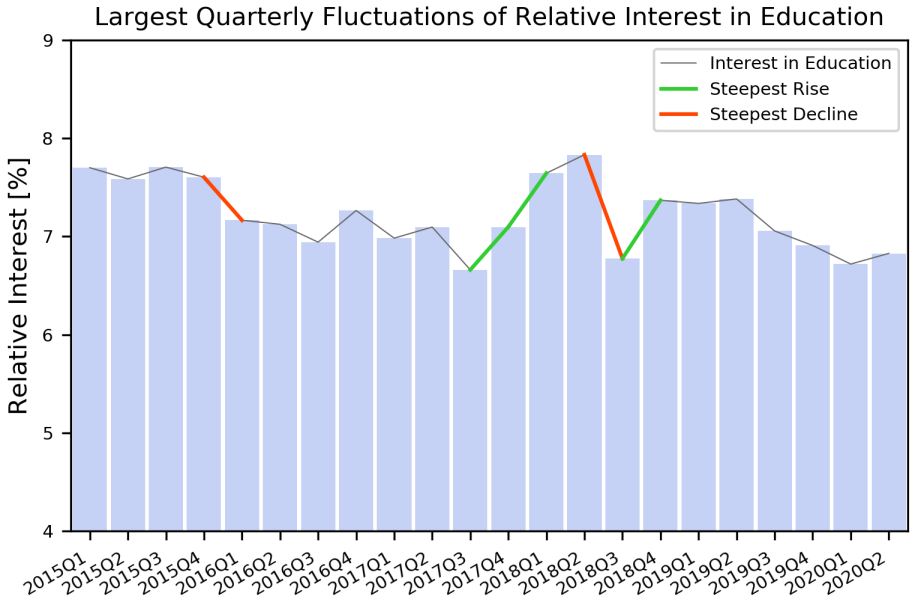


Figure 15. Largest fluctuations of interest in education since 2015

2. Second research question: ***When were the largest fluctuations in the public's interest in education?*** A visual inspection of the plot produced by PIEDI (see Figure 15) reveals that the quarters with the steepest decline in public interest in education were *2016Q1* and *2018Q3*. The steepest rise in public interest in education issues were *2017Q4*, *2018Q1* and *2018Q4*. This list of five quarters with the largest fluctuations of interest in education answers the second question of this dissertation.

3. Third research question: ***Looking at the steepest declines observed, which national issues displaced the public interest in education?*** Based on the answers to the first two questions, I knew that the quarters in which the interest in education plummeted since 2015 were *2016Q1* and *2018Q3*, but I still needed to know which issues of national interest displaced *education* in these quarters.

A human-performed topic labeling process was conducted. Each topic was labeled based on the terms with the highest load in the *Topic Loading Table*. For example, the *topic # 8* was labeled *politics* based on its highest-load terms: *trump*, *clinton*, *president*, *democrats*, *republicans*, *party*, *candidate*. Similarly, the *topic # 13* was labeled *The Environment* based on the terms that topic had with the highest loads: *water*, *plant*, *flint*, *EPA* and *tree*. The topic I labeled *Education* (*topic # 12*) contained the terms: *student*, *school*, *teacher*, *district*, *education*, *community*, *parent*, *child*, *university*, *kid*, *college* and *campus* with the highest loads.

Using the labels I previously assigned, PIEDI produced a report showing the topics that rose during the steepest declines in education (see Figure 16). After applying a threshold of 0.50 (see vertical dashed line) it can be observed that in both quarters *Politics* was the most influential topic.

During the education decline of 2016Q1. During *2016Q1* the issues that notably grew in interest and may have caused the interest in education to drop were the ones previously labeled *Politics*, with a change of 0.856, *The Environment*, with a change of 0.47, *International Trade*, with a change of 0.357 and *International Affairs*, with a change of 0.346 (see Figure 16).

During the education decline of 2018Q3. During *2016Q3* the issues that notably grew in interest and may have caused the interest in education to drop were the ones previously labeled *Politics*, with a change of 0.675, *Elections*, with a change of 0.437, *Local Issues*, with a change of 0.321 and *Healthcare*, with a change of 0.29 (see Figure 16).

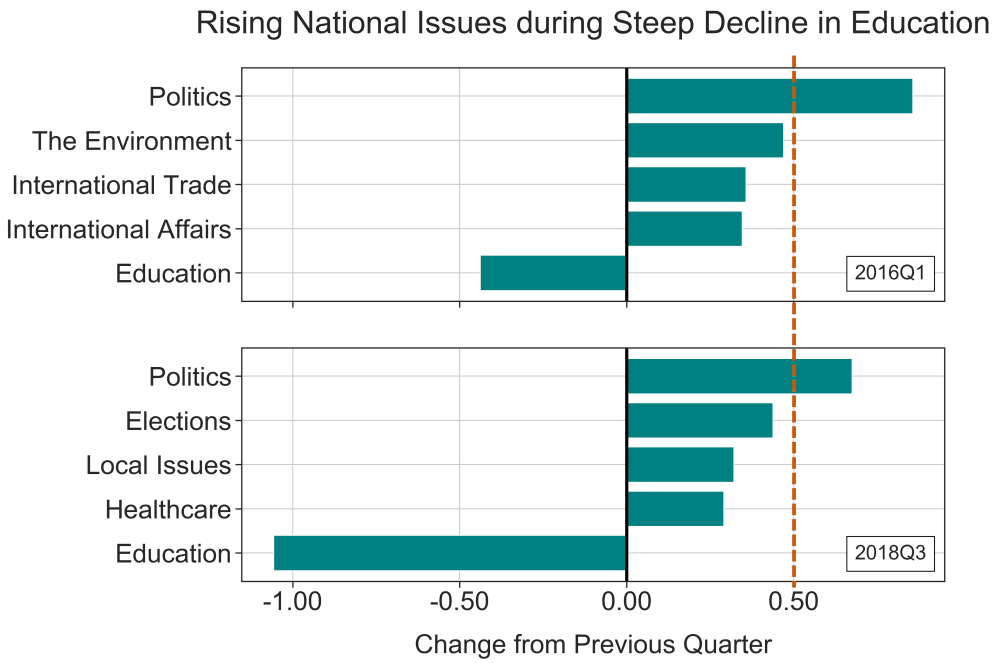


Figure 16. Rising national issues during steep decline of interest in education

The time period *2018Q3*, in which *Education* faced a major drop (a change of −1.056), seems to have been mainly influenced by the mid-term elections of 2018, since the most prominent topics that surged were all related to politics. As can be seen in Figure 16, the topics labeled *Politics*, *Elections* and *Local Issues* made fluctuated upward significantly that quarter. It is possible that *Healthcare* is also related to the political platforms of the candidates, which could explain its rise alongside with the rest of the political issues. If *Politics*, *Elections* and *Local Issues* would have summed up into one group, it would have shown a very significant increase in slope, which would have explained the drop in *Education*.

Partial answer to the third research question. Utilizing a change threshold of 0.5, the partial answer to the third question is that politics issues were likely to be associated with the decline in interest in education issues in *2016Q1*, and in *2018Q3*. The human-performed thematic analysis conducted in the next stage helped understand in detail which specific issues in politics were prominent these quarters.

4. Fourth research question: ***What were the specific education issues that caused the interest in education to surge?*** When events caused interest in education to rise, what were the specific education subtopics that may have caused these surges? To answer that question, I reprogrammed PIEDI further to conduct an additional topic modeling procedure to first discover the education subtopics and then represent them in a time series plot (see Figure 17).

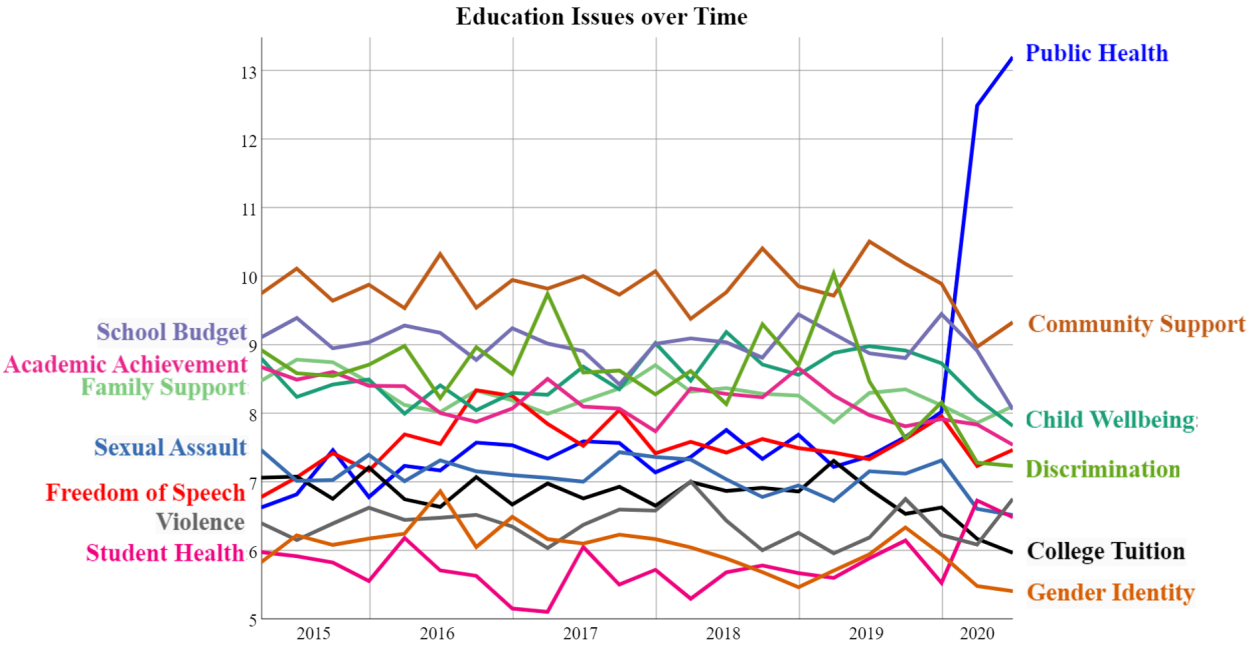


Figure 17. Education issues over time

This new capability of PIEDI of producing subtopics from *Education* (called *education subtopic discovery process*) was programmed by running a new topic modeling process over a subset of the original 30,000 sample of articles. The subset contained the 1,200 articles most related to education from the original sample. This *education subtopics discovery process* created a new set of structures specifically focused on education articles (e.g., *Education Quarterly atf Table*, *Education Topic Loading Table*). Similar to what I have done previously, I experimented with different number of subtopics to select the

optimal number to utilize (based on the groups obtained on each topic discovered) and decided to select a total of 13 education subtopics.

Similar to what I have done to answer to the first three questions, the labels of the education subtopics were assigned based on the terms with the highest loads in the *Education Topic Loading Table*. The topic *Education* was now divided into the following 13 subtopics: *Family Support*, *Public Health*, *Freedom of Speech*, *College Tuition*, *Sexual Assault*, *Student Health*, *Community Support*, *Violence*, *Child Wellbeing*, *Gender Identity*, *School Budget*, *Academic Achievement* and *Discrimination*.

The 13 education subtopics discovered are shown in Figure 17, and their fluctuation in interest can be observed visually (see Figure 17). As can be seen, *Public Health* subtopic of education increased dramatically in *2020Q1*, which is understandable given the fact that this was the time of the outbreak of the 2020 coronavirus pandemic, COVID-19, which dramatically impacted schools and colleges. It is also easy to see the impact the pandemic had on other areas of public interest in education. For example, the interest in the education subtopic *Gender Identity* dropped in that same quarter. All *Education* subtopics, such as *Violence*, *Sexual Assault* or *College Tuition*, could consequently be tracked over time to analyze their fluctuations since 2015.

During the education surge of 2017Q4. During *2017Q4* the *Education* issues that notably grew in interest and may have caused the interest in *Education* to increase were the ones previously labeled *Child Wellbeing*, with a change of 0.673, *School Budget*, with a change of 0.593, *Community Support*, with a change of 0.344 and *Family Support*, with a change of 0.341 (see Figure 18).

During the education surge of 2018Q1. During *2018Q1* the *Education* issues that notably grew in interest and may have caused the interest in *Education* to increase were the ones previously labeled *Academic Achievement*, with a change of 0.626, *Violence*, with a change of 0.427, *College Tuition*, with a change of 0.345 and *Discrimination*, with a change of 0.342 (see Figure 18).

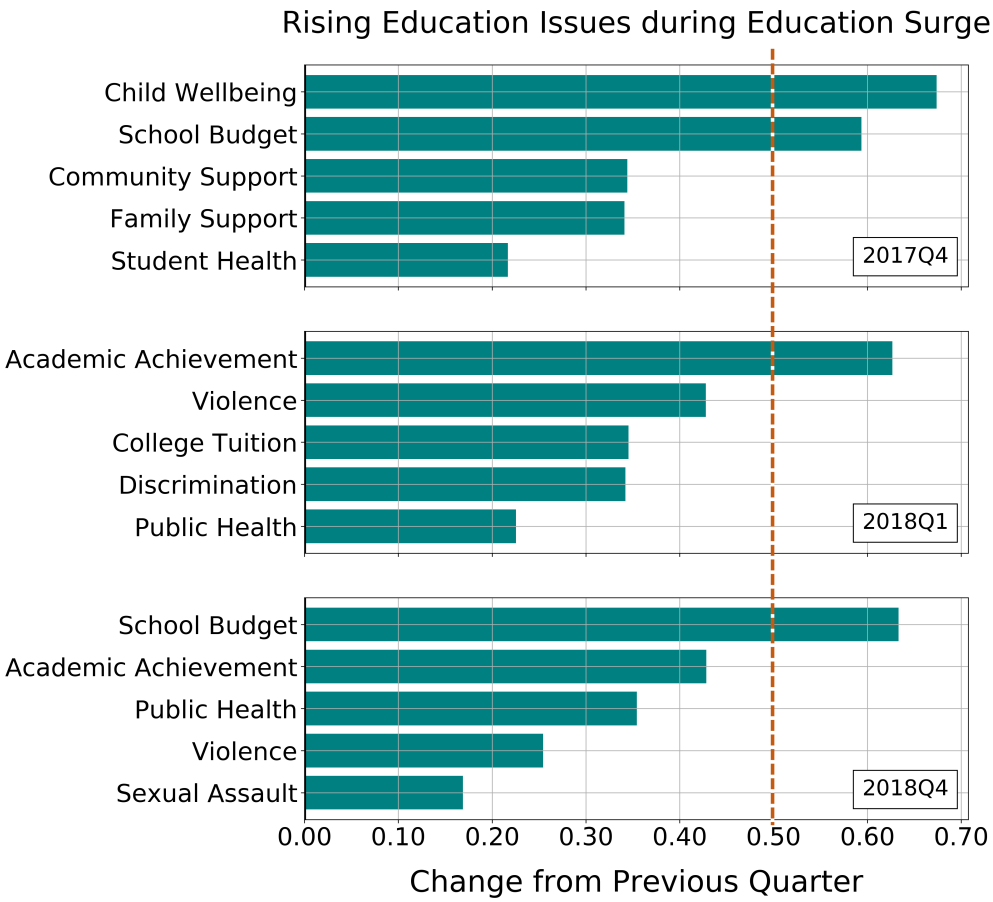


Figure 18. Rising education issues during surges of interest in education

During the education surge of 2018Q4. During *2018Q4* the *Education* issues that notably grew in interest and may have caused the interest in *Education* to increase were the ones previously labeled *School Budget*, with a change of 0.633, *Academic Achievement*, with a change of 0.428, *Public Health*, with a change of 0.354 and *Violence*, with a change of 0.254 (see Figure 18).

Partial answer to the fourth research question. The objective of the fourth question was to analyze which of the education subtopics grew in interest during the quarters of education surge, which were *2017Q4*, *2018Q1* and *2018Q4* (see Figure 15 in page 78). To answer this question, PIEDI produced a new plot for these quarters showing the education subtopics that grew in public interest during such quarters (see Figure 18).

Utilizing a change threshold of 0.5, the partial answer to the fourth question is that the education subtopics with increased public interest in *2017Q4* were *Child Wellbeing* and *School Budget*, the most prominent issue in *2018Q1* was *Academic Achievement* , and the issue with the highest growth in public interest in *2018Q4* was *School Budget*. The human-performed thematic analysis conducted in the next stage helped understand in detail which specific issues in child wellbeeing, academic achievement and budget were prominent these quarters.

Summary of Instrument Deployment

The PIEDI instrument was utilized to answer the first two questions of this dissertation. When new data becomes available, for example, next year, the instrument can be executed to read the new articles and produce a new time series of the fluctuation of interest.

The instrument provides the researcher with two outputs a) a time series of the fluctuations of interest and quarterly cohorts of documents. These two outputs are utilized by the researcher to conduct a human-performed thematic analysis

(C) Final Analysis

This stage includes the thematic analysis (TA) of the annotated cohorts of documents produced by PIEDI based on the quarters obtained in the previous phase. The objective of the human-performed qualitative analysis conducted in this stage is not to perform a comprehensive meta-analysis of the related articles, but to obtain the necessary information to better understand the specific events that occurred during these times of large fluctuations.

The quarterly cohorts of annotated documents for a specific *quarter* and *topic* are the documents published in that *quarter* with the highest scores when their row in the *atf Table* is dot-multiplied by the *topic* column of the *Topic Loading Table*. A cutoff of 20 documents per relevant topic and quarter was applied. Cohorts of documents were

produced only on the *topics* and *quarters* resulting from the quantitative section of the third research question of this study: *Politics* in *2018Q3* and in *2018Q1*. This association is not exclusive, an article may be relevant to more than one topic (e.g. *Education* and *Politics*). For a list of documents included in these cohorts, see *Appendix C*.

Similarly, the quarterly cohorts of annotated documents for a specific *quarter* and *education subtopic* are the documents published in that *quarter* with the highest scores when their row in the *atf Table* is dot-multiplied by the *subtopic* column of the *Education Topic Loading Table*. A cutoff of 20 documents per relevant subtopic and quarter was also applied here. Cohorts of documents were produced only on the *subtopics* and *quarters* resulting from the quantitative section of the fourth research question of this study: *Child Wellbeing* in *2017Q4*, *Academic Achievement* in *2018Q1*, and *School Budget* in *2018Q4*. For a list of documents included in these cohorts, see *Appendix C*.

Quarters with largest sudden declines in interest in education. According to the answers to the second research question of this study (see Figure 16 in page 80), the quarters with the largest plunge of interest in education were *2016Q1* and *2018Q3*. During these quarters, there were indeed topics of national interest that surged considerably. The TA conducted here was able to identify the characteristics of these national issues and synthesize the content written in the news about them.

2016Q1: Politics topic. The common theme in the politics-related news articles published in *2016Q1* was the discussion of the political platforms of the presidential candidates, with particular emphasis on controversial ideas by from candidates. In addition to reporting victories in local primaries, many of the articles focused on the anti-establishment premise, and more specifically, some articles (e.g., Gerson, 2016; Krauthammer, 2016a, 2016b)) expressed doubts regarding the core beliefs and political philosophy of then primary candidate Donald Trump. Notable issues mentioned were tariffs on Chinese imports, universal healthcare, racism and immigration policies. Additionally, that quarter was dominated by news related to manufacturing jobs and foreign policy.

After reading the articles related to this topic, I could conclude that the label *Politics* clearly represents the written content during this period.

2018Q3: Politics and elections topic. Although the topic modeling algorithm separated these two topics, a similar theme can be observed in them, with a slight difference in the subjects they cover. While the articles related to the topics labeled *Politics* mostly discussed the political views of the parties and their representatives, the news related to *elections* were more focused on caucus, the election engine and numerical predictions. In these articles, economy and immigration occupied the majority of the headlines (e.g., Rice, 2018). Also discussed in the news were the approval rates for President Donald Trump, and family separations at the border, and the federal government shutting over budget concerns (e.g., Miller, 2018).

The cohort of documents of *2018Q3* showed that pre-midterm elections could be very influential when it comes to public interest and may have been a factor that affected the drop in education during that quarter.

Quarters with largest surges of interest in education. Based on the quantitative partial results of the fourth question of this dissertation, *2017Q4*, *2018Q1* and *2018Q4* were the periods in which the public interest in education surged (see Figure 18 in page 83). In addition, through a topic modeling process performed over the education issues, a total of 13 topics and their time series since 2015 were uncovered. The subtopics of greater surge in each quarter were analyzed using a human-performed qualitative analysis. The findings are reported below.

2017Q4: Child wellbeing - education subtopic. The articles published in *2017Q4* related to the topic of child wellbeing covered a large variety of issues faced by students at risk. Many authors published works related to bullying and the associated mental scars. Reporters mentioned the mental health issues caused by bullying, including anxiety and depression, which could even lead to suicide, while some others discussed the causes associated with harmful behaviors among children (e.g., Putnam, 2017).

Mental health problems were not uncommon in the news of *2017Q4*. Some news articles even concentrated on the parents of children with mental health problems and their denial attitude towards the issues (e.g., Saker, 2017). Additional issues covered in the news were concussion, gender issues and the associated stigma, domestic violence, poverty, and, in particular, the potential trauma to children of immigrants facing deportation.

All articles I read referred mental health in some form. Some of the reporters concentrated on bullying, anxiety and depression, while others focused on the support for students with mental health issues. Based on the articles, I concluded that during this quarter, the most prominent issue in education was *Mental Health*.

2018Q1: Academic achievement - education subtopic. The news articles related to this topic in *2018Q1* mainly addressed the multiple struggles faced by students at all levels, from preschool to college, the impact these issues have on academic achievement, and the steps taken by school administrators and legislators to measure and improve the lives of the students.

The use of seclusion and restraint as systems to correct behavior was widely discussed in the news, mentioning, among other problems, the abuse exercised at certain schools, and the racial disparities in suspension rates (e.g., McLaren, 2018). Some articles focus on corporal punishment which was apparently being used at a higher rate on students with disabilities (e.g., Pignolet, 2018).

Lastly, prevalent in *2018Q1* was the subject of student achievement and the gaps between the different socioeconomic populations. Some of the articles discuss the methods used to measure academic achievement and the accountability levels of the local school officials (e.g., Walters, 2018), while many others discuss legislation topics related to student success (e.g., Luiz, 2018).

2018Q4: School budget - education subtopic. Budget constraints, legislation and elections for school boards nation-wide were the most prevalent topics discussed in the news the last quarter of 2018. Among other topics, some reporters

addressed propositions for new taxes to support school needs (e.g., Mansfield, 2018), while some others concentrated on the financial crisis faced by many school districts (e.g., Rangel, 2018), or government funding of schools (e.g., Shupe, 2018). There were articles specifically talking about education funding reforms (e.g., Sun, 2018).

The second most talked subject discussed in the news relates to elections for the State Board of Education, the different districts and their candidates and how they plan to address the budget deficit of the schools (e.g., Erbacher, 2018; Hall, 2018).

Third research question: Looking at the steepest declines observed, which national issues displaced the public interest in education? Based on the thematic analysis conducted in this stage, I was able to complete the answer to the third question of this study:

- The most significant topic that may have contributed to a drop of interest in education during *2016Q1* was *Politics*.
- The most significant issues that may have contributed to a drop of interest in education during *2018Q3* were related to politics and pre-midterm elections.

Fourth research question: What were the specific education issues that caused the interest in education to surge? Based on the thematic analysis conducted in this stage, I was able to complete the answer to the fourth question:

- The topic that contributed to the surge of interest in education in *2017Q4* were *Child Welbeing*, in particular *Mental Health*, and *School Budget*
- The most salient issues that may have contributed to the surge of interest in education in *2018Q1* were related to racial disparity and academic gaps.
- The interest during *2018Q4* was dominated by education news related school budget.

Closing Remarks

This dissertation is categorized in two ways. First, it is a mixed methods design based on the Exploratory Design for Instrument Development by Creswell and Plano Clark (2006). Second, this study is considered a phenomenological longitudinal exploratory mixed methods design based on the four research questions it attempts to answer. The first

two questions produced a quantitative exploratory result, while the last two questions were answered through a phenomenological study of the issues reported by newspapers.

This study contained three major stages: the first one was the development of the software instrument, the second one was the utilization of the instrument created (to fully answer the first two questions and partially answer the last two), and the third one was the interpretation of the selected news articles.

The dissertation required the development of a computation-intensive software program able to read tens of thousands of news articles, conduct statistical modeling analysis on the text extracted, and then produce reliability and validity reports by correlating the results of the instrument with a gold standard.

Once the software was ready, employing it was straight-forward and did not require much human interaction as the instrument performed most of the tasks in an automated way. The output of the instrument included a time series of the fluctuations of interest in education, a list of quarters in which the interest in education fluctuated the most, and the quarterly cohorts of relevant articles for further analysis.

The final stage was to conduct a human-performed qualitative phenomenological study of the quarterly cohorts of relevant articles pertaining to the quarters in which the fluctuations of interest in education were larger. This final stage completed the answers to the third and fourth questions of this dissertation.

CHAPTER 5: DISCUSSION

The objective of this research was not to replace polls analyzing the public interest in education, but to augment their capabilities through a technological advancement study. First, by identifying the quarters in which largest fluctuations of interest in education occurred, I was able to analyze the factors that might have influenced (or were associated with) these fluctuations. Second, by reading the news related to each period of large fluctuation I was able to describe the influencing factors by synthesizing text written by journalists during the period.

A software instrument was developed for this study to collect news articles related to national issues in the US since 2015. The instrument crawled tens of thousands of web pages, extracted the terms from the text and analyzed the frequency of appearance of these terms. Only articles maintained by consumer-driven professional gatekeepers from newspapers recognized for archival by the Library of Congress were analyzed. In addition to gathering the text, the tool conducted an automated qualitative coding and grouping of terms to uncover latent topics in the news related to issues in the US. The instrument was deemed reliable after obtaining a high correlation between its outputs when it was executed using three different samples of news articles. The instrument was further considered a valid tool to describe variations in public interest in education over time after the fluctuations of interest in education produced by the tool (based on news exposure) were found highly correlated with the results of a gold standard poll published by Reuters, which monitors public opinion through online surveys.

The share of a particular issue among the interest in all national issues is what matters, rather than the absolute number of people that actually consider that issue a national priority. Any related entity that acts upon priorities (e.g., the press, the Congress) has limited capacity to address issues, and a specific issue that grows significantly in importance could displace other issues, even while they grow in absolute numbers. For

example, if a war starts, the topic *The Economy* could decline despite becoming an even bigger issue in absolute terms. For this reason, the national polls report the percentage of interest in each issue, and consequently, this study also utilized *relative interest* as the metric of choice. Therefore, any reference to interest in issues (including education) in this study implies *relative interest*.

The analysis of the text of online news published since 2015 resulted in a time series of quarterly values representing the interest in education issues over time. The largest fluctuations were identified using both a visual and a statistical analysis. The quarters in which the largest fluctuations of interest in education occurred were analyzed further to determine what specific issues may have influenced such fluctuations in education.

It was found that during the quarters of largest decline of interest in education issues related to politics and national elections rose considerably. On the other hand, surges of public interest in education were found to be associated with the *Education* subtopics of child wellbeing, including mental health, bullying, and effects of domestic violence on students. One of the most prominent issues that increased in importance in the eyes of the public when interest in education climbed was academic achievement (or student success).

Interpretation of the Findings

Using PIEDI, a software instrument I developed for this study to conduct automated qualitative text analysis, I obtained 20 unnamed topics (representing issues in the U.S.), each containing a list of related terms ordered by a score, or *load*, that represented the level of significance the terms had within that topic. For example, *Topic #14* had the following significant terms: *patient*, *hospital*, *doctor*, *drug*, *marijuana*, *treatment*, *study*, *child*, *addiction*, *care* and *medication*, while *Topic #18* contained these significant terms: *mexico*, *immigrant*, *border*, *cuba*, *migrant*, *refugee*, *trump*, *united states*, *immigration*, *country* and *child*. I assigned a label to each topic based on the most significant words they represented. For example, *Topic #14* was labeled *Healthcare*, and *Topic #18* was labeled *Immigration*. After analyzing the 20 topics obtained and assigning

a label to each, the final topic labels were: *International Affairs*, *Local Issues*, *Science & Technology*, *Russia*, *Public Health*, *International Trade*, *U.S. Mail*, *Morality*, *Politics*, *Sports*, *Gender Identity*, *Other*, *Education*, *The Environment*, *Healthcare*, *Elections*, *Crime*, *The Economy*, *Immigration* and *Legislation*.

For each of the topics labeled, I obtained a value, representing the percentage of media coverage for that topic each quarter. For example, if the value for *The Environment* in *2019Q3* was 7.4, this means that 7.4% of the content in news articles which covered issues in the US that quarter was about environmental issues.

The first two questions of this study were answered with the help of Figure 15 on page 78. This plot shows, in addition to the blue bars representing the *Relative Interest in Education* for each quarter, the fluctuations of interest quarter by quarter. The bold red lines show the largest drops, and the bold green lines the largest surges. As can be observed in the plot, there appears to be some seasonality of interest with annual cycles, dropping every year in the third quarter. This drop of interest in education can be seen during the third quarters of 2016, 2017, 2018 and 2019. A possible explanation for this is the decreased public interest in education during the summer break when schools are out.

For the periods when *Education* plunged, it was important to understand which other national issues surged, likely displacing education in public interest as compared to the preceding quarter. To achieve this, I retrieved the national issues that grew the most during the quarters of education sudden drops, which where *2016Q1* and *2018Q3*. I found that *Politics* was the most prominent topic that surged during these quarters of *Education* decline. While this finding does not prove that *Politics* was the cause of *Education* plunge, it seems that in times of election the public pays more attention to other political issues, somewhat displacing the interest in education. Issues related to education are not salient during election time, likely because they are not a subject of large controversy in the otherwise polarized landscape of the US politics.

The *Education Relative Time Series*, expressing the coverage of education relative to all topics analyzed, was produced, tested for validity, and then used as a barometer for monitoring the public interest in education over time. Its further division into subtopics enabled a deeper analysis of the most prominent issues in *Education*. A total of 13 subtopics were identified (see Figure 17 on page 81), labeled manually based on their most significant terms, and ranked based on the public interest in each subtopic each quarter.

The two education subtopics with the highest continuous rank since 2015, labeled *Community Support* and *School Budget*, focused on education issues in the community and budget considerations for schools. Following these, *Discrimination* shows peaks of public interest in 2017Q1 and 2019Q1. The label *Discrimination* was assigned based on my interpretation of some its most significant terms: *black*, *white*, *poverty*, *color*, *race* and *white student*, and others.

While without the thematic analysis of the articles related to the surges it may not be possible to positively determine what may have caused them, the surge of 2017Q1 may have been related to the *Equity in Higher Education Act* of *Chapter 888* approved on September 30, 2016, which specifies diversity requirements for higher-education institutions. Other sudden increases in public interest observed (which would require a detailed thematic analysis to determine their causes) were the surge of *Freedom of Speech* in 2017Q3 and of *Gender Identity* in 2016Q2.

The subtopic *Public Health* was labeled based on terms, such as *COVID*, *pandemic*, *virus*, *mask*, *china*, *fall*, *coronavirus* and *international student*. Undoubtedly, the Coronavirus pandemic outbreak of 2020 (starting in the US in 2020Q1) caused an unprecedented disruption to the education system, with classes transitioned to online delivery, student dormitories closed, and all in-person activities (including sporting events) suspended. The impact of the pandemic was so acute, that (with the single exception of the related subtopic labeled *Student Health*) the interest in the rest of the subtopics of education plunged significantly at that time.

While dividing the national issues into topics was not the objective of this study, as part of my research a total of 20 latent topics representing problems in the US emerged (see Figure 6 on page 65): *International Affairs*, *Local Issues*, *Science & Technology*, *Russia*, *Public Health*, *International Trade*, *U.S. Mail*, *Morality*, *Politics*, *Sports*, *Gender Identity*, *Other*, *Education*, *The Environment*, *Healthcare*, *Elections*, *Crime*, *Economy*, *Immigration* and *Legislation*. These 20 topics (obtained by the PIEDI instrument) were matched with the list of response options that Reuters/Ipsos have for their poll question (see Figure 6 on page 65).

Theoretical Implications

The conceptual framework of this dissertation started with the paper by Nehoran and Nehoran (2020), which described the steps to follow when creating an instrument with the objective of inferring the public opinion through textual content analysis of news published and monitored by professional consumer-driven gatekeepers. By following the relevant parts of the guidelines, I created the PIEDI instrument, which was programmed to analyze the text of news articles to understand the public interest in issues in the US, particularly as they relate to education. I found a significant correlation between the changes in *Education* interest levels obtained by PIEDI and the corresponding variations in the published gold standard I chose for this study (Reuters/Ipsos). This correlation matched the claims made by Nehoran and Nehoran (2020), who maintained that certain data regarding public opinion could be inferred from the news. Furthermore, the literature review revealed multiple studies which showed that due to the symbiotic influence that exists between the media and the public, there is a correlation between them (e.g., McCombs & Shaw, 1972). When a topic gains prominence in the news it attracts the public attention and then is read (and shared) more, which influences the news outlets to publish more related articles (McCombs, 2004).

As described in the background of this study, professional consumer-driven gatekeepers do not make decisions in isolation, but they play in a field of influencing

factors, one of which is their financial motivation, based mainly on advertisements (Vu, 2014). The more popular an article (or topic) becomes, the more revenue they generate from advertisements. The gatekeepers monitor the performance of their articles to understand the preferences of their audiences through metrics, such as number of unique visitors, time spent reading the articles, number of shares, and comments made (Tandoc, 2014; Vu, 2014; Welbers et al., 2016).

The literature review of this study covers the use of the term *salience*, referring to the prominence gatekeepers attribute to a topic, measured mainly by frequency and level of exposure. Similarly, *salience* in the public agenda refers to the importance given to certain topics by the public, measured through polls, clicks, shares, "likes", time spent on web pages, and more. As explained in the literature review, in Communication Theory, *influence* happens when the salience is transferred from one agent to the other. When the media increases the salience of a particular topic, the public reacts and becomes more interested in that topic, which affects the salience in the public agenda. Similarly, when the public becomes very interested in a topic, people click more on related articles, which encourages gatekeepers to increase the frequency of the topic in the news.

Fake News, Misinformation and Hidden Agendas

While readership metrics are not the only factors influencing the media, the fact that politicians, wealthy individuals, and interest groups are constantly attempting to advance their agendas through the media (and many of them are successful in doing so), does not impact the claim used in this study, that there is a correlation between the salience of an issue in the media and in the public interest (regardless of the origin). In fact, if the media is influenced by a certain interest group to increase the prominence of a particular topic (unrelated to other motives involved), this will influence the readers to read more about it, which will encourage the media to publish more related articles later on. The topic gaining salience (in the media and public agendas) might even be a scandal, misinformation, or fake news, but this does not change the claim I maintain in this

research, that there is a correlation between the media salience and the public interest salience of any particular topic. This matches my findings in Chapter 4 of this study (in this case, specifically related to *Education*).

Contribution of This Study

This study tracked the public interest in the topic *Education* since 2015 to learn how it fluctuated quarter by quarter. The second component of the conceptual framework by Burstein (2003) (see Figure 1 on page 18), emphasizes the influence public opinion has on policymaking. The results of this study will help policymakers working on education issues analyze historical trends to a) evaluate whether increase of interest in certain topics in education have triggered new legislation in the past and b) comprehend the impact made by specific education-related regulations on public opinion. More informed policymakers are more likely to make data-driven decisions when they propose education-related laws.

Methodological Implications

Monitoring the fluctuations of interest in education over the years is not new. In fact, many polling agencies (such as Gallup and Reuters/Ipsos) have tracked *Education* as part of a larger poll, asking people to name the most burning issue or issues. For consistency of results and comparability, these polling agencies kept the exact same question (text and response options) from the day they introduced it. Gallup's (open-ended) question text is: *"What do you think is the most important problem facing the country today?"*, while Reuters/Ipsos' is: *"In your opinion, what is the most important problem facing the US today? (Select from below or write in)"*.

While the fixed question, which ran for decades, provided a good way of monitoring certain issues over the years (e.g., *Education*, *Immigration*), it has two main problems: a) it cannot account for societal changes or new trends, and thus is not able to adapt to these changes, and b) it asks about just one, single burning topic, which is more sensitive to sensational news or sporadic events.

Adaptation to Societal Changes

Gallup's poll contains an open-ended question, which does not dictate specific issues people can mention, so keeping their same question over time allows the respondents to incorporate new, contemporary issues that may arise over time. In contrast, the poll from Reuters/Ipsos contains the following list of options: *The Economy*, *Immigration*, *Energy issues*, *Crime*, *Unemployment/Lack of Jobs*, *Terrorism*, *Morality*, *The Environment*, *War/Foreign Conflicts*, *Healthcare*, *Education*, *I don't know*, and *Other*. By using the same options since 2012, Reuters/Ipsos frames the train of thought of the respondents around these specific issues, ignoring new critical ones that might have emerged since then. Although some respondents probably use the *Other* category and write in an additional issue, I believe that giving the respondents a list of predetermined options discourages many of them from thinking of, or taking the time to write in, other possible responses. Based on my findings (see Figure 6 on page 65), some notable categories that might be missing from Reuters/Ipsos list of choices are: *Sports*, *Gender Identity*, and most importantly *Science and Technology*. It is possible that these categories are missing from their list of options because these issues were not as important in 2012 (when Reuters/Ipsos' poll started). In today's age of technology, I believe that it would have been prudent to include a *Science and Technology* category, as it is very likely that cybersecurity, privacy and social media issues are all important problems that people care about.

Misleading Indicators on Important Issues

By asking the public the "*what is the...issue/problem*" question, the researcher assumes that the public cares more (or exclusively) about one issue, which may not be true. I believe that forcing the respondents to select a single most important problem may create bias against lower ranked issues. For example, people may have divided opinions about the *most important* current issue (some may believe it is *The Economy*, while others may mention *Immigration* or *Public Health*) but they might all agree on the top three issues. Furthermore, the majority of the respondents might believe, for example, that

Education is second in priority, but *Education* would not be reported as a high-priority issue unless enough people mention it as their top priority.

Contribution of This Study

While the aggregated answers of an MIP poll question provide valuable information about the priorities assigned by the public, a single term, such as *The Economy*, *The Environment* or *Healthcare* does not provide enough information describing what is the specific issue (or issues) that people care about. For example, when respondents mark *Healthcare*, they may be worried about healthcare costs, about the ability of the system to take care of their health, or about related legislation. Similarly, when *Education* is selected in the poll, it may mean that people are worried about the quality of education, about the school environment, or about college tuition. This study complements the results of education-related polls by providing a way to associate news articles to the topics the way they are grouped by the polling agency. By using such association, a human-performed thematic analysis can be performed to synthesize the articles related to the topics and better understand what people care about.

This study supplements surveys (which are not easily adaptable to societal changes) by providing the power to add new topics at a later time without affecting the capability to track and maintain consistency of topics over time. For example, while *Science & Technology* is not included in the Reuters/Ipsos poll as one of the topics, because it was not deemed important enough to justify a separate category in 2012, the methodology used in this study allows adding it now as a topic, and track it retroactively without affecting the rest of the topics. Another topic which also grew in public interest since 2012 is *Gender Identity*.

Policy and Practical Implications

Public opinion matters. Politicians and decision-makers (including policymakers) consult the metrics of public interest at decision time, particularly if they are seeking

reelection. Based on the background section of this study and the literature review, these metrics can come from polls, social media posts, or news articles. The more salient the topics are, the more likely they will attract the attention of decision-makers.

Contribution of This Study

The instrument developed in this study provides new information that helps stakeholders related to *Education* concentrate on what is important to the public.

By providing a list of 13 subtopics of *Education* and how each of them fluctuated since 2015, this study helps stakeholders target specific subtopics representing issues of interest in *Education*, (e.g., *School Budget*, *Sexual Assault in Colleges*, *Gender Identity*, or *Student Health*), take note of how these issues fluctuate, and try to find ways to solve them by a) proposing new legislation, b) increasing awareness to these issues to encourage solutions, c) starting new related research, or d) dedicating funds to help address the issues.

Policymakers. Policymakers related to *Education* now have a new way to understand which subtopics are the ones people care most about. Even just a visual examination of Figure 17 on page 81 provides policymakers with a descriptive picture of the *Education* issues in the US in unprecedented detail. Policymakers can now use this to track certain topics of interest (e.g., *Child Wellbeing*, *Freedom of Speech*). They can then try to understand the fluctuations and propose legislation to help address the issues.

Journalists and gatekeepers. Newspaper employees reporting on *Education* can use the results of this study to complement the dashboards they consult regularly and have more information about the *Education* issues the public cares about most. For example, Figure 17 on page 81 shows that *Community Support*, *Discrimination in School* and *School Budget* have consistently been the top three issues people care about, followed by *Child Wellbeing*. These are good topics to be explored by journalists for their upcoming articles.

Scholars. Scholars have conducted research in *Education* for more than a century, targeting issues related to curriculum, standardized tests, learning styles, discrimination, or racism in schools, just to name a few. This research provides an additional mechanism that

informs educational scholars which issues are important to the public and how particular issues of the scholar's interest have fluctuated over time. Furthermore, it can facilitate the thematic analysis to investigate the causes associated with the fluctuations of interest.

Sponsors. Philanthropists are continually searching for projects to sponsor. Each of them uses their own methods to select which causes to promote either for the welfare of others, or to advance research on topics they care about. The list of *Education* issues produced in this study increases sponsors' awareness on issues that people care about, which can help them prioritize the allocation of donations to proposed projects.

Assumptions and Limitations

The PIEDI instrument was validated with the news articles published since 2015 by the newspapers recognized for archiving by the Library of Congress. Using the validated instrument with earlier data, or with news articles from other sources may require additional validity tests, without which the results may be inaccurate. Thus, the PIEDI instrument is limited to news articles from the newspapers in this particular list that were published 2015 or later.

PIEDI is limited to English text and does not support other languages. To cover newspapers in all languages that appear on the list of the Library of Congress, separate versions of the instrument need to be developed.

This study is limited to covering the interest level of those who read the news online. Despite the fact that gatekeepers still use traditional methods to measure readership interest in printed articles (in addition to the online metrics), their outreach to paper-edition readers is limited. It is possible that the opinion of the people who only read printed news is slightly different than the conventional online news consumer. In addition, this study does not cover the opinion of the individuals who do not read the news at all.

The gold standard used to validate PIEDI uses internet polls to collect public opinion. It is possible that the results of another data collection methods (e.g. phone interview, focus groups) may result in different opinions from the ones completing online surveys.

This study analyzed the fluctuation of interest in issues in the US, and cannot be generalized to other countries, nor narrowed down to smaller geographies such as US states or counties.

This study assumed that the majority of the newspapers recognized for archival by the Library of Congress are consumer driven, that is, are operating to satisfy their readership (e.g., in order to increase page views and gain financial benefit).

An additional limitation faced by researchers who conduct time series analyses relates to granularity. Specifically, I had to consider the best level of aggregation afforded by the data available online as of 2020. Higher levels of granularity (such as using bi-weekly or monthly data points) would necessarily be based on fewer articles, potentially yielding less accurate estimates of interest. Lower levels of granularity (such as using semi-annual data points) would involve more articles and potentially more accurate estimates. However, as compared to the quarters utilized in this study, there would be fewer pairs of bivariate points available on which to base the correlation which was used to evaluate the concurrent validity of the instrument. Furthermore, the findings as to how interest in education fluctuates, where it drops and where it surges may be impacted. Thus, to balance statistical power and accuracy vs. validity quality, my findings are based on the quarterly data points.

Recommendations for Future Research

This study provided a new way to separate the education issues in the US into subcategories in a scientific manner, and identified their fluctuation of interest over the years since 2015: *Family Support*, *Public Health*, *Freedom of Speech*, *College Tuition*, *Sexual Assault*, *Student Health*, *Community Support*, *Violence*, *Child Wellbeing*, *Gender Identity*,

School Budget, *Academic Achievement* and *Discrimination*. A full understanding of the reasons for the fluctuations in each of these education categories may be a topic for future research.

This research divides the topic *Education* into 13 categories, each depicting one of the most salient subtopics in education (e.g., *Gender Identity in Schools*, *Violence in Colleges*, *Sexual Assault in Colleges*). However, other subdivisions may be made by future research to focus on a different segmentation of education. For example, the topic *Education* could be divided into categories, such as *K-12* and *Higher Education*. Additionally, the public interest level in each of these education subtopics may have been influenced by local or national laws proposed or may have influenced legislators to consider passing new laws. The analysis of possible correlation between the fluctuations of these education categories and the related laws proposed are potential topics for future research.

This study found that politics and elections surged during periods of largest drops of interest in education. However, future research may further analyze the data to determine if politics and national elections are a regular cause of interest drop, and to what extent the elections play a role in distracting the public attention away from education.

A combination of direct and partial matches between PIEDI and Reuters/Ipsos' MIP question enables a potential future thematic analysis of each of Reuters/Ipsos categories through the content of the related news articles associated with them. For example, a thematic analysis of the articles related to PIEDI's topic labeled *The Economy* will help better understand what the public means when they mark *The Economy* or *Unemployment* on Reuters/Ipsos' poll.

This technological advancement study included the creation of a software instrument programmed to conduct many of the labor-intense tasks of qualitative analysis in an automated way. Future research could extend these capabilities beyond the education topic and focus on other topics covered in the news and the polls.

Among the six natural language processing levels, as defined by Crowston et al. (2012) (morphological, lexical, syntactic, semantic, discourse, and pragmatic), this study analyzes text up to the lexical level. The lexical level identifies the part of speech, marking the difference in significance of a word (e.g. "bear") between a noun and a verb. This study does not understand how the order of words influences their meaning, which is considered the syntactic level. Future research could apply the text analysis at the syntactic level and enable an automated detection of concepts from the text. For example in the current study, the phrase *the budget cut is necessary to avoid improper spending of taxpayer dollars* leads to the following isolated terms *budget cut*, *improper spending* and *taxpayer dollars*, while a syntactic level study could join these terms to report that the *budget cut* was related to *improper spending of taxpayer dollars*.

As mentioned in the limitations, the findings from this study are likely to depend upon the granularity of the time series analyses. I expect that the fast growing demand for online news across all U.S. regions will cause a large increase in news agencies that make their articles available online. At the same time, since a growing demand for news articles is translated to potential revenue, this will also encourage consumer-driven gatekeepers to increase the frequency of their online publications. I believe that in a few years, this increase in online availability could allow a future researcher to conduct a similar study with a finer granularity (e.g., bi-monthly, monthly, or even bi-weekly). Since it is likely that articles that are now only in print version will later become available online, future research could also potentially replicate this study for the same time period with higher granularity. This will allow an evaluation of how the reliability and validity of the instrument, as well as other findings, are impacted by the choice of granularity.

Conclusion

This study focused on understanding the national issues associated with the plunges of public interest in education and the education issues that rose considerably when a surge in education was observed.

By developing a software instrument programmed to collect the text of tens of thousands of online documents and to perform the typically time-consuming task of qualitative analysis theme discovery (coding and grouping into topics) in an automated way, I was able to analyze large amounts of information (more than $50,000$ documents) that would have been time-prohibitive otherwise. The time series of interest in education issues produced in this study, was validated against the gold standard Reuters/Ipsos poll, and thus considered a valid barometer to measure the changes in public interest in education in the US.

This technological advancement study not only showcases how the public interest can be inferred from the text of published news articles, but also that by combining a) automated text retrieval and analysis tools with b) automated document classification procedures and c) human-performed thematic analyses, it is possible for researchers to discover topics in the text of a very large repository of documents, associate the most relevant documents to these topics, and ultimately read these relevant documents to interpret the meaning of the written text.

The results showed that issues related to particularly politics and elections grew in interest during the times of the steepest declines in the interest in education. In smaller proportions, international trade, foreign conflicts, housing issues and substance abuse all seemed to attract the public attention and possibly displaced the importance of education in the public eye.

Additionally, it was found that particular education problems rose considerably in public interest during the quarters of surge of interest in education, among them school budget, academic achievement and child wellbeing (particularly focusing on mental health issues, bullying and domestic violence).

Lastly, this study contributed to the body of knowledge by providing a new way to separate the issues into subcategories in a scientific manner through latent subtopics that comprised *Education* as a topic of national interest in the US, and how these subtopics

fluctuated over time since 2015: *Family Support*, *Public Health*, *Freedom of Speech*, *College Tuition*, *Sexual Assault*, *Student Health*, *Community Support*, *Violence*, *Child Wellbeing*, *Gender Identity*, *School Budget*, *Academic Achievement* and *Discrimination*.

REFERENCES

References marked with an asterisk (*) indicate studies included in the research synthesis

Allahyari, M., Pouriyeh, S., Assefi, M., Safaei, S., Trippe, E. D., Gutierrez, J. B., & Kochut, K. (2017). A brief survey of text mining: Classification, clustering and extraction techniques, In *Kdd bigdas*, Halifax, Canada. http://arxiv.org/abs/1707.02919

Back to school statistics (tech. rep.). (2019). National Center for Education Statistics. https://nces.ed.gov/fastfacts/display.asp?id=372

Bannigan, K., & Watson, R. (2009). Reliability and validity in a nutshell. *Journal of Clinical Nursing*, *18*(23), 3237–3243. https://doi.org/10.1111/j.1365-2702.2009.02939.x

Bardes, B., & Oldendick, R. (2012). Public opinion and American democracy (Fourth), In *Public opinion. measuring the american mind* (Fourth). Rowman & Littlefield Publishers.

Bartee, R. D., & Brown, C. (2007). Insight into the different forms of capital, *312*, 1–47. https://www.jstor.org/stable/42979737

Blanchett Neheli, N. (2018). News by Numbers: The evolution of analytics in journalism. *Digital Journalism*, *6*(8), 1041–1051. https://doi.org/10.1080/21670811.2018.1504626

Blumer, H. (1953). The mass, the public, and the public opinion. *Reader in public opinion and communication*, 43–49.

Bourdieu, P. (1986). The forms of capital. https://www.marxists.org/reference/subject/philosophy/works/fr/bourdieu-forms-capital.htm

Burstein, P. (2003). The impact of public opinion on public policy: A review and an agenda. *Political Research Quarterly*, *56*(1), 29–40. https://doi.org/10.1177%2F106591290305600103

Campbell, D. E. (2006). What is education's impact on civic and social engagement? *Measuring the Effects of Education on Health and Civic Engagement*, (March), 1–387. http://www.oecd.org/education/innovation-education/37425694.pdf

Christin, A. (2018). Counting clicks: Quantification and variation in web journalism in the United States and France. *American Journal of Sociology*, *123*(5), 1382–1415. https://doi.org/10.1086/696137

Core Political Data Ipsos Poll Conducted for Reuters (tech. rep.). (2016). Ipsos Public Affairs. http://big.assets.huffingtonpost.com/2016ReutersTrackingCorePolitical6222016.pdf

Core Political Data Ipsos Poll Conducted for Thomson Reuters (tech. rep.). (2020). Ipsos Public Affairs. https://www.ipsos.com/sites/default/files/ct/news/documents/2020-04/2020_reuters_tracking_-_core_political_coronavirus_tracker_04_08_2020.pdf

Creswell, J. W., & Plano Clark, V. L. (2006). Choosing a mixed methods design (second), In *Designing and conducting mixed methods research* (second). SAGE Publications.

Crowston, K., Allen, E. E., & Heckman, R. (2012). Using natural language processing technology for qualitative data analysis. *International Journal of Social Research Methodology*, *15*(6), 523–543. https://doi.org/10.1080/13645579.2011.625764

*Erbacher, M. (2018). EVSC School Board: Jean Webb, District 1 candidate. https://www.courierpress.com/story/news/politics/elections/2018/10/18/evsc-school-board-jean-webb-district-1-candidate-elections/1682528002/

Fahmi, U., Wibowo, C. P., & Yudanto, F. (2018). Hybrid method in revealing facts behind texts: A combination of text mining and qualitative approach, Atlantis Press. https://doi.org/10.2991/aapa-18.2018.14

Feldman, R., & Sanger, J. (2009). Introduction to text mining, In *The text mining handbook*. https://doi.org/10.1017/cbo9780511546914.002

*Gerson, M. (2016). Republican Party might not survive a Trump nomination. https://www.cjonline.com/article/20160107/OPINION/301079767

Glynn, C., Herbst, S., Lindeman, M., O'Keefe, G. J., & Shapiro, R. Y. (2016a). Dimensions of public opinion, In *Public opinion*. Taylor & Francis Group.

Glynn, C., Herbst, S., Lindeman, M., O'Keefe, G. J., & Shapiro, R. Y. (2016b). Methods for studying public opinion, In *Public opinion*. Taylor & Francis Group.

Glynn, C., Herbst, S., Lindeman, M., O'Keefe, G. J., & Shapiro, R. Y. (2016c). The meaning of public opinion, In *Public opinion*. Taylor & Francis Group.

Groshek, J., & Groshek, M. C. (2013). Agenda trending: Reciprocity and the predictive capacity of social networking sites in intermedia agenda setting across topics over time. *Media and Communication*, *1*(1), 15–27. https://doi.org/10.12924/mac2013.01010015

Guetterman, T. C., Chang, T., DeJonckheere, M., Basu, T., Scruggs, E., & Vydiswaran, V. G. (2018). Augmenting aualitative text analysis with natural language processing: Methodological study. *Journal of medical Internet research*, *20*(6), e231. https://doi.org/10.2196/jmir.9702

*Hall, S. (2018). Kentucky midterm election 2018: State House candidates guide. https://www.courier-journal.com/story/news/politics/elections/kentucky/2018/10/23/kentucky-voter-guide-state-house-general-election-candidates-2018/1444941002/

Hodgkiss, M. (2017). Newsworthiness guidelines for a socially responsible press: Aligning definitions at the intersection of journalism, ethics, and the law. https://scholarworks.gsu.edu/cgi/viewcontent.cgi?article=1091&context=communication_diss

Hynds, E. C. (1981). Newspaper education coverage has been expanded, improved. *Newspaper Research Journal*, *2*(3), 70–76. https://doi.org/10.1177/073953298100200309

Jackson, C. (2020). *Public poll findings and methodology Reuters / Ipsos : Trump's approval holds steady at 38 % , driven largely by partisan divides* (tech. rep.). Gallup Publications. https://www.ipsos.com/en-us/reutersipsos-core-political-presidential-approval-tracker-07082020

Jacobson, H. K. (1973). Needed improvements in education news coverage as perceived by media and education gatekeepers. *Journal of Educational Research*, *66*(6), 274–278. https://doi.org/10.1080/00220671.1973.10884479

Jæger, M. M. (2010). Does cultural capital really affect academic achievement? *Sociology of Education*, *84*(0001), 281–298. https://doi.org/10.1177/0038040711417010

Janasik, N., Honkela, T., & Bruun, H. (2009). Text mining in qualitative research: Application of an unsupervised learning method. *Organizational Research Methods*, *12*(3), 436–460. https://doi.org/10.1177/1094428108317202

Jiang, Y., Qi, Y., Wang, W. K., Bent, B., Avram, R., Olgin, J., & Dunn, J. (2020). EventDTW: An improved dynamic time warping algorithm for aligning biomedical signals of nonuniform sampling frequencies. *Sensors (Switzerland)*, *20*(9). https://doi.org/10.3390/s20092700

Jones, J., & Saad, L. (2020). *Gallup news service. July wave 1* (tech. rep. June 2019). Gallup Publications.

Kim, Y. (2013). Contingent factors of agenda-setting effects: How need for orientation, issue obtrusiveness, and message tone influence issue salience and attitude strength (T. Johnson, Ed.). In T. Johnson (Ed.), *Agenda setting in a 2.0 world: New agendas in communication*. Taylor & Francis Group. https://doi.org/10.4324/9780203735879

*Krauthammer, C. (2016a). After Iowa, Republicans' options redefined. https://www.dailynews.com/2016/02/04/after-iowa-republicans-options-redefined-charles-krauthammer/

*Krauthammer, C. (2016b). Trump election would be death blow to conservatism – East Bay Times. https://www.eastbaytimes.com/2016/02/05/charles-krauthammer-trump-election-would-be-death-blow-to-conservatism/

Krippendorff, K. (2019). *Content analysis: An introduction to its methodology* (4th). SAGE Publications, Inc. https://doi.org/10.2307/2288384

Li, W., & McCallum, A. (2006). Pachinko allocation: DAG-structured mixture models of topic correlations, In *Twenty-third international conference on machine learning*. https://doi.org/10.1145/1143844.1143917

Library of Congress. (2020). https://www.loc.gov/rr/news/ncr_list.php

Lippmann, W. (1922). The world outside and the pictures in our heads, In *Public opinion*. Harcourt, Brace; Co.

*Luiz, J. (2018). BCSD board candidates discuss major education issues at forum. https://www.bakersfield.com/news/bcsd-board-candidates-discuss-major-education-issues-at-forum/article_0e6c61f6-28c5-11e8-8284-c3269e453ee3.html

MacGregor, P. (2007). Tracking the online audience, metric data start a subtle revolution' journalism studies. *Journalism Studies*, *8*(2), 280–298. http://dx.doi.org/10.1080/14616700601148879

*Mansfield, A. (2018). Gardner mayor imposes hiring freeze after voters reject tax override for schools. https://www.telegram.com/news/20181107/gardner-mayor-imposes-hiring-freeze-after-voters-reject-tax-override-for-schools

Martin, J. A. (2013). Agenda setting, elections, and the impact of information technology (T. Johnson, Ed.). In T. Johnson (Ed.), *Agenda setting in a 2.0 world: New agendas in communication*. Taylor & Francis Group. https://doi.org/10.4324/9780203735879

Marvasti, A. B. (2019). Qualitative content analysis: A novice's perspective. *Qualitative social Research*, *20*(3). https://doi.org/10.17169/fqs-20.3.3387

Matthes, J. (2005). The need for orientation towards news media: Revising and validating a classic concept. *International Journal of Public Opinion Research*, *18*(4), 422–444. https://doi.org/10.1093/ijpor/edh118

McCombs, M. E. (1968). The influence of mass communication in politics. *Today's Speech*, *16*(4), 31–34. https://doi.org/10.1080/01463376809385506

McCombs, M. E. (2004). *Setting the agenda: The mass media and public opinion* (G. Mesch, Ed.).

McCombs, M. E. (2005). A look at agenda-setting: Past, present and future. *Journalism Studies*, *6*(4), 543–557. https://doi.org/10.1080/14616700500250438

McCombs, M. E., & Shaw, D. L. (1972). The agenda-setting function of mass media. *Oxford University Press*, *36*(2), 176–187. https://doi.org/10.1086/267990

McCombs, M. E., Shaw, D. L., & Weaver, D. H. (2014). New Directions in Agenda-Setting Theory and Research. *Mass Communication and Society*, *17*(6), 781–802. https://doi.org/10.1080/15205436.2014.964871

McCombs, M. E., & Zhu, J.-h. (1995). Capacity , diversity , and volatility of the public agenda : Trends from 1954 to 1994. *American Association for Public Opinion Research*, *59*(4), 495–525. https://doi.org/10.1086/269491

*McLaren, M. (2018). Could JCPS be taken over by Kentucky? Here's what we know. https://www.courier-journal.com/story/news/education/2018/02/05/jcps-kentucky-state-takeover-questions/300992002/

*Miller, Z. (2018). Trump says 'no problem' shutting government, dismaying GOP. https://www.adn.com/nation-world/2018/07/31/trump-says-no-problem-shutting-government-dismaying-gop/

Nehoran, D., & Nehoran, Y. I. (2020). Inferential public opinion research: A development model for document analysis instruments investigating public opinion through professionally edited media. *Unpublished manuscript*.

O'Connor, C., & Joffe, H. (2020). Intercoder reliability in qualitative research: Debates and practical guidelines. *International Journal of Qualitative Methods*, *19*, 1–13. https://doi.org/10.1177/1609406919899220

O'Keefe, G. J. (1978). Newspaper versus television reliance and political disaffection. *Theory and Methodology Division*.

*Pignolet, J. (2018). Corporal punishment: Students with disabilities receive it more in Tennessee, report shows. https://www.commercialappeal.com/story/news/education/2018/03/15/corporal-punishment-schools-tennessee-disabilities-students/427988002/

Popkin, S. l. (1991). The effects of education, In *The reasoning voter: Communication and persuasion in presidential campaigns*. The University of Chicago Press.

*Putnam, C. (2017). Why kids bully: causes and ways to prevent it. https://www.democratandchronicle.com/story/lifestyle/2017/12/07/what-causes-some-kids-hurtful-and-how-turn-them-around/920612001/

Raich, M., Müller, J., & Abfalter, D. (2014). Hybrid analysis of textual data: Grounding managerial decisions on intertwined qualitative and quantitative analysis. *Management Decision*, *52*(4), 737–754. https://doi.org/10.1108/MD-03-2012-0247

*Rangel, M. (2018). Opinion: New fiscal crisis requires Oakland schools change. https://www.mercurynews.com/2018/11/01/opinion-new-financial-crisis-requires-oakland-schools-finally-change/

Reuters Poll Explorer. (n.d.). https://polling.reuters.com/poll/SC8

Reuters Poll Explorer Methodology. (n.d.). Retrieved October 3, 2020, from https://staging.polling.reuters.com/methodology

Reuters/Ipsos/UVA Center for Politics State Poll: Tennessee (tech. rep.). (2018). Ipsos Public Affairs. https://www.ipsos.com/sites/default/files/ct/news/documents/2018-10/topline_-_tennessee_-_final_10.17.18.pdf

*Rice, P. (2018). The biggest issue in the 6th District: Trump. https://www.news-journalonline.com/opinion/20180805/pat-rice-biggest-issue-in-6th-district-trump

Rose, J., & Lennerholt, C. (2017). Low cost text mining as a strategy for qualitative researchers. *Electronic Journal of Business Research Methods*, *15*(1), 2–16. http://www.ejbrm.com/volume15/issue1/p2

*Saker, A. (2017). What's top concern for Cincinnati parents? Their kids' mental health. https://www.cincinnati.com/story/news/2017/10/11/no-1-worry-cincinnati-parents/751503001/

Sattikar, A. A., & Kulkarni, R. V. (2012). Natural language processing for content analysis in social networking. *International Journal of Engineering Inventions*, *1*(4), 2278–7461. https://pdfs.semanticscholar.org/ce7c/8fe3409b3805f948e1f0f93617c265182180.pdf

Senin, P. (2008). Dynamic Time Warping Algorithm Review. *Science*, *2007*(December), 1–23. https://doi.org/10.1109/IEMBS.2007.4353810

Serban, S. (2015). On the origin of the gatekeeping theory and its application to journalism. *Annals of Spiru Haret University*, *16*(2), 12–24.

Shaw, D. L., & McCombs, M. E. (1977). The emergence of American political issues: The agenda-setting function of the press. *West Publishing Co.*, 211.

*Shupe, B. (2018). More government spending isn't the answer for Delaware. https://www.delawareonline.com/story/opinion/contributors/2018/10/15/more-government-spending-isnt-answer-delaware/1646083002/

Stauffer, A., Oliff, P., Foard, C., Thiess, R., Reese, M., Samms, B., Bryant, M., Pontari, L., Spence, E., & Einsiedler, J. (2017). *Two decades of change in federal and state higher education funding* (tech. rep.).

Stern, S., Livan, G., & Smith, R. E. (2020). A network perspective on intermedia agenda-setting. *Applied Network Science*, *5*(1), arXiv 2002.05971. https://doi.org/10.1007/s41109-020-00272-4

Strömbäck, J., Karlsson, M., & Hopmann, D. N. (2012). Determinants of news content: Comparing journalists' perceptions of the normative and actual impact of different event properties when deciding what's news. *Journalism Studies*, *13*(5-6), 718–728. https://doi.org/10.1080/1461670X.2012.664321

Sullivan, A. (2001). Cultural capital and educational attainment. *Sociology*, *35*(04), 893–912. https://doi.org/10.1017/S0038038501008938

*Sun, L. (2018). For students' sake, fix education funding law. https://www.lowellsun.com/2018/11/28/editorial-for-students-sake-fix-education-funding-law/

Tandoc, E. C. (2014). Journalism is twerking? How web analytics is changing the process of gatekeeping. *New Media and Society*, *16*(4), 559–575. https://doi.org/10.1177/1461444814530541

Tran, H. (2013). Online agenda setting: A new frontier for theory development (T. Johnson, Ed.). In T. Johnson (Ed.), *Agenda setting in a 2.0 world: New agendas in communication*. Taylor & Francis Group. https://doi.org/10.4324/9780203735879

Vaismoradi, M., & Snelgrove, S. (2019). Theme in qualitative content analysis and thematic analysis. *Forum Qualitative Sozialforschung*, *20*(3). https://doi.org/10.17169/fqs-20.3.3376

Vayansky, I., & Kumar, S. A. (2020). A review of topic modeling methods. *Information Systems*. https://doi.org/10.1016/j.is.2020.101582

Vu, H. T. (2014). The online audience as gatekeeper: The influence of reader metrics on news editorial selection. *Journalism*, *15*(8), 1094–1110. https://doi.org/10.1177/1464884913504259

Wade, S., & Schramm, W. (1969). The mass media as sources of public affairs, science, and health knowledge. *Public Opinion Quarterly*, *33*(2), 197. https://doi.org/10.1086/267691

Wallace, J. (2018). Modelling contemporary gatekeeping: The rise of individuals, algorithms and platforms in digital news dissemination. *Digital Journalism*, *6*(3), 274–293. https://doi.org/10.1080/21670811.2017.1343648

*Walters, D. (2018). The yin and yang of California's school crisis. https://www.eastbaytimes.com/2018/01/21/walters-the-yin-and-yang-of-californias-school-crisis/

Wanta, W., & Ghanem, S. (2000). Effects of agenda-setting (J. Bryant & R. Carveth, Eds.). In J. Bryant & R. Carveth (Eds.), *Meta-analysis of media effects*. Erlbaum, Mahwah.

Wanta, W., & Alkazemi, M. F. (2017). Agenda-setting: History and research tradition. *The International Encyclopedia of Media Effects*, (March 2017), 1–14. https://doi.org/10.1002/9781118783764.wbieme0030

Weaver, D. H. (1977). Political issues and voter need for orientation (M. E. McCombs & D. L. Shaw, Eds.). In M. E. McCombs & D. L. Shaw (Eds.), *In the emergence of american political issues*.

Welbers, K., Van Atteveldt, W., Kleinnijenhuis, J., Ruigrok, N., & Schaper, J. (2016). News selection criteria in the digital age: Professional norms versus online audience metrics. *Journalism*, *17*(8), 1037–1053. https://doi.org/10.1177/1464884915595474

Wells, R. (2008). The effects of social and cultural capital on student persistence. *Community College Review*, *36*(1), 25–46. https://doi.org/10.1177/0091552108319604

West, D. M., Whitehurst, G. J. R., & Dionne, E. J. (2009). *Invisible: 1.4 percent coverage for education is not enough* (tech. rep.). Governance Studies at Brookins.

Yamamoto, M., Nah, S., & Chung, D. S. (2017). U.S. newspaper editors' ratings of social media as influential news sources. *International Journal of Communication*, *11*, 684–700. https://doi.org//10.1086/696137

Yu, C. H., Jannasch-Pennell, A., & DiGangi, S. (2011). Compatibility between text mining and qualitative research in the perspectives of grounded theory, content analysis, and reliability. *Qualitative Report*, *16*(3), 730–744. https://nsuworks.nova.edu/tqr/vol16/iss3/6/

Yuan, J., Lin, Q., Zhang, W., & Wang, Z. (2019). Locally slope-based dynamic time warping for time series classification, In *International conference on information and knowledge management, proceedings*. https://doi.org/10.1145/3357384.3357917

APPENDIX A: INSTRUMENT IMPLEMENTATION REPORT

Source

1. **Source Selection**: All newspapers recognized for archival by the Library of Congress constitute the list of selected news agencies for this study
2. **Source Qualification Criteria**: Among the newspapers selected in previous step, the following ones were excluded:
    - **Language**: Newspapers written in a language different from English
    - **Online availability**: Newspapers without online presence
    - **Subscription**: Newspapers to which I did not have access due to subscription requirements
    - **Duplicated Domains**: Newspapers that shared the domain with another ones
3. **Sample Size**: The number of news agencies selected for this study was 258 (See Appendix B) for the full list

Search

1. **Engine**: The Bing engine by Microsoft was utilized for this study
2. **Query**: The query used was "*issues in America*"
3. **Document Collection Procedure**:
    - **Crawler**: A self-developed software web crawler was utilized to run the query "*issues in America*" and extract all documents that were published by the selected newspapers
    - **Results**: The crawler extracted up to 3 pages of 10 web search results and all related queries provided by Bing
4. **Sample Size (number of articles retrieved from the web)**: A total of 113,457 articles were collected using the Bing search engine to search articles related to issues in America published by the selected newspapers
5. **Sample Composition**: The selected repository of documents included all documents that resulted from running the query on the Bing Engine using the crawler to extract all articles published by the selected newspapers since 2015.

Articles

1. **Article Qualification Criteria**: Due to the inefficiency of the Bing engine (and general web search issues related to promotion of specific web pages, problems that are out of the scope of this dissertation), not all the articles retrieved were related to issues in America. In addition, only articles monitored by gatekeepers are relevant to this study. Among the publications selected, the following news articles were excluded:

    - **Articles unrelated to issues**: Articles containing less than 5% negative terms were excluded. Obituaries, Horoscopes and Transcripts were excluded.
    - **Letters to the Editor**: Letters to the editor were excluded
    - **Article size**: Articles with less than 300 words or more than 10,000 words were excluded
    - **Article terms**: Articles that did not contain the terms within the $10,000$ most popular ones in the repository were excluded from the study

2. **Sample Size (number of selected articles)**: After applying the article qualification criteria, the total number of articles was 51,262.

Text in Articles

Published news contain, in addition to the text, a large number of words that are irrelevant to the content of the articles. They include text from advertisements, promotions, invitation to events, and text from related news. I created a software text analyzer able to retrieve the text from the documents and exclude irrelevant terms.

1. **Text Qualification**: The terms excluded include:

    - Names of months and weekdays
    - Names of months and weekdays
    - Dates and times
    - Numbers (except when combined with letters, such as in the case of COVID-19)
    - paragraphs with links or email addresses

2. **Text Modification**s:

    - **Lower Case**: as it is common on natural language processing methods, all text was transformed to lower case (including proper names)
    - **Singular**: as it is common on natural language processing methods, all text was transformed to singular form

Terms

Terms are formed when two or more words are joined and treated as a group of words, because together they provide a different meaning that each of such words separately. Some examples of multi-word terms are *financial aid*, *congress representative*, or *college board*.

1. **Term rule**: For this study, I defined a term as being either a single noun/adjective, or multiple adjacent nouns/adjectives, including proper nouns (e.g., *interdisciplinary school program*)

2. **Library Utilized**: NLTK: The Natural Language Toolkit, by E. Loper, and S. Bird (2002)

Topics

1. **Library Utilized**: The tomotopy library of Python was utilized to extract the topics from the text

2. **Number of topics**: I decided to extract a total of 20 topics for national issues and 13 topics for education issues.

APPENDIX B: SELECTED NEWSPAPERS

The following list contains the newspapers selected for this study. Only the main newspaper is listed when multiple newspapers use the same domain.

#	State	City	Newspaper Name
1)	Alabama	Birmingham	Birmingham News
2)	Alabama	Montgomery	Montgomery Advertiser
3)	Alaska	Anchorage	Anchorage daily news
4)	Alaska	Fairbanks	Fairbanks Daily News-Miner
5)	Alaska	Juneau	Juneau Empire
6)	Arizona	Flagstaff	Arizona Daily Sun
7)	Arizona	Phoenix	Arizona Republic
8)	Arizona	Tucson	Arizona Daily Star
9)	Arizona	Yuma	Yuma Daily Sun
10)	Arkansas	Little Rock	Arkansas Democrat-Gazette
11)	California	Bakersfield	Bakersfield Californian
12)	California	Brisbane	Hellenic Journal
13)	California	Long Beach	Press-Telegram
14)	California	Los Angeles	Asbarez
15)	California	Los Angeles	Daily News
16)	California	Los Angeles	Los Angeles Sentinel
17)	California	Los Angeles	Los Angeles Times
18)	California	Oakland	Oakland Tribune
19)	California	Riverside	Press Enterprise
20)	California	San Diego	San Diego Union-Tribune
21)	California	San Francisco	Philippine News
22)	California	San Francisco	San Francisco Chronicle

#	State	City	Newspaper Name
23)	California	San Francisco	Wall Street Journal
24)	California	San Jose	Mercury News
25)	California	San Jose	Tribuna Portuguesa
26)	California	Santa Ana	Orange County Register
27)	Colorado	Colorado Springs	Gazette
28)	Colorado	Denver	Denver Post
29)	Colorado	Pueblo	Pueblo Chieftain
30)	Connecticut	Bridgeport	Connecticut Post
31)	Connecticut	Danbury	News-Times
32)	Connecticut	Hartford	Hartford Courant
33)	Connecticut	New Haven	New Haven Register
34)	Delaware	Dover	Delaware and Maryland state news
35)	Delaware	Wilmington	News Journal
36)	District of Columbia	Washington	International Iran times
37)	District of Columbia	Washington	USA Today
38)	District of Columbia	Washington	Washington Informer
39)	District of Columbia	Washington	Washington Post
40)	District of Columbia	Washington	Washington Times
41)	Florida	Daytona	Daytona Beach news-journal
42)	Florida	Ft. Lauderdale	South Florida Sun-Sentinel
43)	Florida	Jacksonville	Florida Times-Union
44)	Florida	Miami	Miami Times
45)	Florida	Orlando	Orlando Sentinel
46)	Florida	Pensacola	Pensacola News Journal
47)	Florida	Sarasota	Herald-Tribune
48)	Florida	St. Petersburg	Tampa Bay times

#	State	City	Newspaper Name
49)	Florida	Tallahassee	Tallahassee Democrat
50)	Florida	West Palm Beach	Palm Beach Post
51)	Georgia	Atlanta	Atlanta Daily World
52)	Georgia	Atlanta	Atlanta Journal-Constitution
53)	Georgia	Augusta	Augusta Chronicle
54)	Georgia	Savannah	Savannah Morning News
55)	Hawaii	Honolulu	Honolulu star-advertiser
56)	Hawaii	Wailuku	Maui News
57)	Idaho	Idaho Falls	Post Register
58)	Idaho	Lewiston	Lewiston Tribune
59)	Idaho	Pocatello	Idaho State Journal
60)	Illinois	Arlington Heights	Daily Herald
61)	Illinois	Bloomington	Pantagraph
62)	Illinois	Chicago	Chicago Defender (daily edition)
63)	Illinois	Chicago	Chicago Sun-Times
64)	Illinois	Chicago	Chicago Tribune
65)	Illinois	Chicago	Draugas
66)	Illinois	Joliet	KSKJ Voice
67)	Illinois	Peoria	Journal Star
68)	Illinois	Springfield	State Journal Register
69)	Indiana	Evansville	Evansville Courier and Press
70)	Indiana	Ft. Wayne	Journal Gazette
71)	Indiana	Ft. Wayne	News and Sentinel
72)	Indiana	Indianapolis	Indianapolis Star
73)	Indiana	Indianapolis	National Jewish Post and Opinion
74)	Indiana	South Bend	South Bend Tribune

#	State	City	Newspaper Name
75)	Iowa	Davenport	Quad-City Times
76)	Iowa	Des Moines	Des Moines Register
77)	Iowa	Dubuque	Telegraph-Herald
78)	Iowa	Mason City	Mason City Globe-Gazette
79)	Iowa	Sioux City	Sioux City Journal
80)	Kansas	Atchison	Atchison Daily Globe
81)	Kansas	Emporia	Emporia Gazette
82)	Kansas	Topeka	Topeka Capital-Journal
83)	Kentucky	Frankfort	State Journal
84)	Kentucky	Lexington	Lexington Herald-Leader
85)	Kentucky	Louisville	Courier-Journal
86)	Kentucky	Owensboro	Messenger-Inquirer
87)	Louisiana	Baton Rouge	Advocate
88)	Louisiana	New Orleans	Advocate
89)	Maine	Augusta	Central Main Sunday
90)	Maine	Bangor	Bangor Daily News
91)	Maine	Lewiston	Sun Journal
92)	Maine	Portland	Portland Press Herald
93)	Maryland	Annapolis	Capital
94)	Maryland	Baltimore	Baltimore Afro-American
95)	Maryland	Baltimore	Sun
96)	Maryland	Frederick	Frederick News-Post
97)	Maryland	Hagerstown	Morning Herald
98)	Maryland	Salisbury	Daily Times
99)	Massachusetts	Boston	Boston Globe
100)	Massachusetts	Boston	Boston Herald

#	State	City	Newspaper Name
101)	Massachusetts	Edgartown	Vineyard Gazette
102)	Massachusetts	Lowell	Sun
103)	Massachusetts	New Bedford	Portuguese Times
104)	Massachusetts	Springfield	Republican
105)	Massachusetts	Watertown	Armenian Mirror-Spectator
106)	Massachusetts	Watertown	Armenian Weekly
107)	Massachusetts	Worcester	Telegram & Gazette
108)	Michigan	Dearborn	Sada al-Watan
109)	Michigan	Detroit	Detroit Free Press
110)	Michigan	Detroit	Detroit Jewish News
111)	Michigan	Detroit	Detroit News
112)	Michigan	Grand Rapids	Grand Rapids Press
113)	Michigan	Lansing	Lansing State Journal
114)	Michigan	Madison Heights	Italian Tribune
115)	Michigan	Marquette	Mining Journal
116)	Minnesota	Duluth	Duluth News-Tribune
117)	Minnesota	Minneapolis	Star Tribune
118)	Minnesota	St. Paul	Pioneer Press
119)	Mississippi	Biloxi-Gulfport	Sun-Herald
120)	Mississippi	Columbus	Commercial Dispatch
121)	Mississippi	Jackson	Clarion-Ledger
122)	Missouri	Jefferson City	Missouri news tribune
123)	Missouri	Kansas City	Kansas City star
124)	Missouri	Springfield	Springfield news-leader
125)	Missouri	St. Louis	St. Louis post-dispatch
126)	Montana	Billings	Billings Gazette

#	State	City	Newspaper Name
127)	Montana	Butte	Montana Standard
128)	Montana	Great Falls	Great Falls Tribune
129)	Montana	Helena	Helena Independent Record
130)	Montana	Pablo	Char-Koosta News
131)	Nebraska	Lincoln	Journal Star
132)	Nebraska	Omaha	World-Herald
133)	Nebraska	Scottsbluff	Star-Herald
134)	Nevada	Carson City	Nevada Appeal
135)	Nevada	Las Vegas	Las Vegas Sun
136)	Nevada	Las Vegas	Review-Journal
137)	Nevada	Reno	Reno Gazette-Journal
138)	New Hampshire	Concord	Concord Monitor
139)	New Hampshire	Manchester	Union Leader
140)	New Jersey	Asbury Park	Asbury Park Press
141)	New Jersey	Atlantic City	Press of Atlantic City
142)	New Jersey	Cherry Hill (Camden)	Courier-Post
143)	New Jersey	Hackensack	Record
144)	New Jersey	Jersey City	Jersey Journal
145)	New Jersey	Newark	Italian Tribune News
146)	New Mexico	Albuquerque	Albuquerque Journal
147)	New Mexico	Farmington	Daily Times
148)	New Mexico	Las Cruces	Las Cruces Sun-News
149)	New Mexico	Santa Fe	Santa Fe New Mexican
150)	New York	Albany	Times-Union
151)	New York	Brooklyn	Hamodia
152)	New York	Brooklyn	Jewish Press

#	State	City	Newspaper Name
153)	New York	Buffalo	Buffalo News
154)	New York	Melville	Newsday (Nassau edition)
155)	New York	Monsey	Yated Ne'eman (English edition)
156)	New York	New York	Daily News
157)	New York	New York	El Diario-La Prensa
158)	New York	New York	Forward (English edition)
159)	New York	New York	Irish Voice
160)	New York	New York	Journal of Commerce
161)	New York	New York	New York Post
162)	New York	New York	New York Times
163)	New York	Rochester	Democrat and Chronicle
164)	New York	Syracuse	Post-Standard
165)	New York	White Plains	Journal News
166)	North Carolina	Asheville	Asheville Citizen-Times
167)	North Carolina	Charlotte	Charlotte Observer
168)	North Carolina	Durham	Herald-Sun
169)	North Carolina	Fayetteville	Fayetteville Observer
170)	North Carolina	Greensboro	Greensboro News & Record
171)	North Carolina	Raleigh	News & Observer
172)	North Carolina	Wilmington	Star-News
173)	North Carolina	Winston-Salem	Winston-Salem Journal
174)	North Dakota	Bismark	Tribune
175)	North Dakota	Fargo	Forum
176)	North Dakota	Grand Forks	Grand Forks Herald
177)	Ohio	Akron	Akron Beacon Journal
178)	Ohio	Cincinnati	American Israelite

#	State	City	Newspaper Name
179)	Ohio	Cincinnati	Cincinnati Enquirer
180)	Ohio	Cleveland	Call and Post
181)	Ohio	Cleveland	Plain Dealer
182)	Ohio	Columbus	Columbus Dispatch
183)	Ohio	Dayton	Dayton Daily News
184)	Oklahoma	Enid	Enid News and Eagle
185)	Oklahoma	Lawton	Constitution
186)	Oklahoma	Oklahoma City	Oklahoman
187)	Oklahoma	Tulsa	Tulsa World
188)	Oregon	Eugene	Register-Guard
189)	Oregon	Portland	Oregonian
190)	Oregon	Portland	Skanner
191)	Oregon	Salem	Statesman-Journal
192)	Pennsylvania	Allentown	Morning-Call
193)	Pennsylvania	Altoona	Altoona Mirror
194)	Pennsylvania	Erie	Erie Daily Times-News
195)	Pennsylvania	Harrisburg	Patriot-News
196)	Pennsylvania	Lancaster	LNP always Lancaster
197)	Pennsylvania	Middletown	Jednota
198)	Pennsylvania	Philadelphia	Philadelphia Tribune
199)	Pennsylvania	Pittsburgh	New Pittsburgh Courier
200)	Pennsylvania	Pittsburgh	Pittsburgh tribune-review
201)	Pennsylvania	Scranton	Times-Tribune
202)	Pennsylvania	York	York Dispatch
203)	Rhode Island	Pawtucket	Times
204)	Rhode Island	Providence	Providence Journal

#	State	City	Newspaper Name
205)	South Carolina	Charleston	Post and Courier
206)	South Carolina	Columbia	State
207)	South Carolina	Greenville	News
208)	South Carolina	Myrtle Beach	Sun-news
209)	South Dakota	Aberdeen	American News
210)	South Dakota	Pierre	CJ, the Capital Journal
211)	South Dakota	Rapid City	Journal
212)	South Dakota	Sioux Falls	Argus Leader
213)	Tennessee	Chattanooga	Chattanooga times free press
214)	Tennessee	Knoxville	Knoxville News Sentinel
215)	Tennessee	Memphis	Commercial Appeal
216)	Tennessee	Nashville	Tennessean
217)	Texas	Austin	American-Statesman
218)	Texas	Corpus Christi	Caller-Times
219)	Texas	Dallas	Morning News
220)	Texas	El Paso	Times
221)	Texas	Ft. Worth	Star Telegram
222)	Texas	Houston	Chronicle
223)	Texas	Lubbock	Avalanche-Journal
224)	Texas	San Antonio	Express-News
225)	Utah	Logan	Herald-Journal
226)	Utah	Ogden	Standard-Examiner
227)	Utah	Salt Lake City	Salt Lake Tribune
228)	Vermont	Barre-Montpelier	Times Argus
229)	Vermont	Burlington	Burlington Free Press
230)	Vermont	Rutland	Rutland Daily Herald

#	State	City	Newspaper Name
231)	Virgin Islands	Charlotte Amalie	Virgin Island Daily News
232)	Virgin Islands	Christiansted	St. Croix Avis
233)	Virginia	Alexandria	Alexandria Gazette Packet
234)	Virginia	Danville	Danville Register & Bee
235)	Virginia	Lynchburg	News & Daily Advance
236)	Virginia	Norfolk	New Journal & Guide
237)	Virginia	Norfolk	Virginian-Pilot
238)	Virginia	Petersburg	Petersburg Progress-Index
239)	Virginia	Richmond	Richmond Times-Dispatch
240)	Virginia	Roanoke	Roanoke Times & World-News
241)	Virginia	Williamsburg	Virginia Gazette
242)	Virginia	Winchester	Winchester Star
243)	Washington	Olympia	Olympian
244)	Washington	Seattle	Seattle Times
245)	Washington	Spokane	Spokesman-Review
246)	Washington	Tacoma	News Tribune
247)	West Virginia	Charleston	Charleston gazette-mail
248)	West Virginia	Fairmont	Times-West Virginian
249)	West Virginia	Huntington	Herald-Dispatch
250)	West Virginia	Parkersburg	Parkersburg News
251)	West Virginia	Wheeling	Intelligencer
252)	Wisconsin	Green Bay	Green Bay Press Gazette
253)	Wisconsin	La Crosse	La Crosse Tribune
254)	Wisconsin	Madison	Wisconsin State Journal
255)	Wisconsin	Milwaukee	Milwaukee Journal-Sentinel
256)	Wyoming	Casper	Casper Star-Tribune

#	State	City	Newspaper Name
257)	Wyoming	Cheyenne	Wyoming Tribune-Eagle
258)	Wyoming	Laramie	Boomerang

APPENDIX C: COHORTS OF ARTICLES

The following list contains the cohorts of articles I read for the human-performed thematic analysis I conducted in this study.

DURING EDUCATION DECLINE

2016Q1: Politics

1) jewishpress.com: Trump Two-State Deal Possible, But Not Makeable [v

2) progress-index.com: Trump and Clinton win in Tri-Cities - News - The P

3) lowellsun.com: Analysis Megyn Kelly schools Trump on Trump U, and

4) heraldtribune.com: Nicholas Kristof Donald the Dangerous - Opinion

5) sltrib.com: Romney calls Trump a fraud, warns his policies wou

6) dailyherald.com: GOP leaders No place for bigotry in the Republican

7) bangordailynews.com: The establishment nonsense

8) dailynews.com: After Iowa, Republicans options redefined Charles

9) jsonline.com: Paul Ryan talks political priorities not Trump i

10) eastbaytimes.com: Charles Krauthammer Trump election would be death

11) mercurynews.com: Charles Krauthammer Trumps problem is he isnt

12) masslive.com: 5 takeaways from the Republican presidential debat

13) delawareonline.com: The establishment nonsense

14) telegram.com: Does Trump have a health care plan Does it matter

15) cjonline.com: Michael Gerson Republican Party might not survive

16) bangordailynews.com: Clintons embrace of Obama holds risks for general

17) app.com: KRAUTHAMMER The fight for the soul of GOP

18) fayobserver.com: Doyle McManus Donald Trump, the candidate of hope

19) arkansasonline.com: Sanders, Trump pick up NH wins

20) news-leader.com: Our Voice Trump unites country in disgust

2018Q3: Politics

1) sj-r.com: Byron York In Washington, a rare relationship with

2) bostonglobe.com: Hillary Clintons 2016 prognosis of puppet Trump pr

3) news-journalonline.com: PAT RICE The biggest issue in the 6th District Tru

4) pjstar.com: Thiessen The lefts contempt is going to re-elect T

5) progress-index.com: Trump urges ouster of senator he blames for derail

6) montgomeryadvertiser.com: The lefts contempt is going to re-elect Trump

7) adn.com: Trump says no problem shutting government, dismayi

8) caller.com: It would be wrong to force Trumps translator to te

9) lasvegassun.com: Kavanaughs views on presidential powers could be

10) bostonglobe.com: Trump France, others hit by terror may face more

11) heraldtribune.com: Leonhardt The question of Trump and money launderi

12) dailyherald.com: In 10th Congressional District race, both Schneide

13) dailyherald.com: Trump says he has no problem shutting down governm

14) telegram.com: Amid criticism over Putin summit, Trump wants seco

15) lowellsun.com: The lefts contempt will re-elect Trump Lowell Sun

16) pjstar.com: A closer look Trump skips some basic rituals of th

17) sj-r.com: A closer look Trump skips some basic rituals of th

18) bostonglobe.com: Fact check Republicans rush to blame Hillary Clint

19) daily-times.com: Trump cant claim the emergency of the tape is fake

20) sj-r.com: Obama issues scathing critique of Trump, politics

2018Q3: Elections

1) cjonline.com: Larry R Bradley Ranked choice voting is the remedy

2) arkansasonline.com: Back to the people - Arkansas Online

3) tallahassee.com: Election 2018 Were helping you prepare to vote wit

4) freep.com: Precincts run out of ballots in Oakland Co; voters

5) journalgazette.net: Kansas GOP gubernatorial primary remains too close

6) dailyherald.com: Kansas GOP gubernatorial primary remains too close

7) tampabay.com: Five things to watch in the final Republican guber

8) courant.com: Time To Open Primaries To Unaffiliated Voters - Ha

9) sun-sentinel.com: Broward has technical glitch, but election smooth;

10) al.com: 25 voters got wrong ballots in runoff for Mobile H

11) orlandosentinel.com: Florida race for US Senate takes center stage with

12) detroitnews.com: Supreme Court keeps ban on Michigan straight-tick

13) heraldtribune.com: Andrew Gillum Can he win - News - Sarasota Herald

14) jacksonville.com: Black activists and community organizers statewide

15) news-journalonline.com: PAT RICE Send me questions for our upcoming debate

16) detroitnews.com: Bankole Dont vote, dont complain - Detroit News

17) progress-index.com: Kobach steps away from duties in contested Kansas

18) news-journalonline.com: PAT RICE The biggest issue in the 6th District Tru

19) courant.com: Editorial The DMV Isnt The Problem - Hartford Cour

20) pnj.com: Editorial Growing environmental problems might dec

DURING EDUCATION SURGE

2017Q4: Child Wellbeing

1) masslive.com: Massachusetts children with mental health problems

2) chicagotribune.com: Families of children with Down syndrome report mos

3) providencejournal.com: Group decries proposed cuts at DCYF - News

4) freep.com: Consider the trauma immigration raids cause for ch

5) news-leader.com: Edwards Why I accepted the Board of Education appo

6) democratandchronicle.com: Why kids bully causes and ways to prevent it

7) cincinnati.com: Whats top concern for Cincinnati parents Their kid

8) augustachronicle.com: Editorial Tie one off - Opinion - The Augusta Chro

9) chicagotribune.com: Yellen says problems of childhood poverty linger -

10) oregonlive.com: 5-year-old autistic child got lost after Portland

11) tallahassee.com: Opinion HB1 not the cure for school bullies or bul

12) chicagotribune.com: Biggest mistakes parents make with their teens - C

13) courierpostonline.com: Kids and concussions Tips for returning to school

14) capitalgazette.com: US is asked to forgive debt of Corinthian Colleges

15) news-journalonline.com: Theos journey A transgender child at war with his

16) dallasnews.com: Fear of deportation causes mental, health issues i

17) sltrib.com: Theos journey A transgender child at war with his

18) tallahassee.com: Stopping domestic violence a community effort

19) arabamericannews.com: For some kids, bullying may not leave lasting ment

20) arkansasonline.com: State takes control of Earle School District; educ

2017Q4: School Budget

1) arkansasonline.com: State takes control of Earle School District; educ

2) jacksonville.com: Duval school budget woes $178 million - News - The

3) freep.com: Recruiting Michigan teachers Wining and dining, in

4) eastbaytimes.com: Opinion Strong leaders needed to fix Oakland schoo

5) chicagotribune.com: Illinois schools have biggest funding gap in natio

6) montgomeryadvertiser.com: Montgomery public schools Charter schools, loss of

7) news-leader.com: Edwards Why I accepted the Board of Education appo

8) tallahassee.com: Opinion HB1 not the cure for school bullies or bul

9) tennessean.com: TNReady Tennessee education chief promises improve

10) cjonline.com: Accidents, other issues make training key for next

11) lowellsun.com: Building maintenance, special ed cited as top Lowe

12) courant.com: Canton Schools Consider Policy Prohibiting Harassm

13) telegram.com: No competition for school bus contracts poses prob

14) telegram.com: Worcesters North High problems have had states att

15) montgomeryadvertiser.com: For Councilman Arch Lee, a district stuck in the m

16) courierpress.com: Ex-judge officially files letter of intent for sea

17) mcall.com: US House passes bills to prop up CHIP through Marc

18) chicagotribune.com: University funding looms as pressure point in Illi

19) adn.com: Sexually active students must be reported to law

20) chicagotribune.com: To improve schools, let teachers take over - Chica

2018Q1: Academic Achievement

1) washingtonpost.com: Public support for gun restrictions has grown to t

2) commercialappeal.com: SCS students We need fewer tests, more learning re

3) news-journalonline.com: B-CU nursing program slapped with probation - News

4) courier-journal.com: As JCPS awaits full audit results, its fixing its

5) nytimes.com: What College Students Really Think About Free Spee

6) commercialappeal.com: Corporal punishment Students with disabilities rec

7) eastbaytimes.com: Walters The yin and yang of Californias school cri

8) washingtontimes.com: Percentage of public schools with resource officer

9) al.com: How one rural Alabama district is closing the gap,

10) courier-journal.com: Could JCPS be taken over by Kentucky Heres what we

11) indystar.com: Indianapolis high schools could start later, but f

12) clarionledger.com: Trump acknowledges National School Choice Week

13) bakersfield.com: BCSD board candidates discuss major education issu

14) jacksonville.com: Duval officials say they received threats at 17 sc

15) nola.com: About 40 percent of New Orleans high school studen

16) masslive.com: 9th-grader arrested at Holyoke High School for pos

17) juneauempire.com: Local students joining nationwide silent protest

18) dispatch.com: Survey Most residents in struggling US areas respe

19) pe.com: Students across US stage school walkouts to protes

20) savannahnow.com: Students get money, mull lawsuits as Savannah Law

2018Q4: School Budget

1) usatoday.com: Supreme Court will review Virginias racially drawn

2) sun-sentinel.com: Broward schools may have used bogus roofing estima

3) courier-journal.com: Indiana election 2018 West Clark School Board vote

4) telegram.com: Gardner mayor imposes hiring freeze after voters r

5) pressherald.com: Five candidates round out roster for Scarborough s

6) chieftain.com: D60 principals support new management for 2 school

7) mercurynews.com: Opinion New fiscal crisis requires Oakland schools

8) omaha.com: Editorial Timm, Neary are strong candidates for th

9) courier-journal.com: Kentucky midterm election 2018 State House candida

10) delawareonline.com: More government spending isnt the answer for Delaw

11) telegram.com: Worcester school officials, frustrated homeschooli

12) statesmanjournal.com: St Edward leaders to fight use of eminent domain b

13) courierpress.com: EVSC School Board Jean Webb, District 1 candidate

14) clarionledger.com: Will states face lawsuits for failing to fund publ

15) lansingstatejournal.com: On the issues - 24th District state Senate candida

16) commercialappeal.com: Deferred maintenance leads to heating problems, at

17) masslive.com: New coalition will push for more education funding

18) courier-journal.com: Indiana election 2018 Your statehouse midterm cand

19) lowellsun.com: EDITORIAL For students sake, fix education funding

20) delawareonline.com: Opinion We need innovative solutions to the school