# AN ANALYSIS ON TRENDS OF RESEARCH TOPICS IN CIVIC EDUCATION USING DYNAMIC TOPIC MODEL

Mr. Poon Thongsai



A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Education in Educational Statistics and Information
Department of Educational Research and Psychology
FACULTY OF EDUCATION
Chulalongkorn University
Academic Year 2022

# การวิเคราะห์แนวโน้มของประเด็นวิจัยทางพลเมืองศึกษาโดยใช้โมเดลไดนามิกทอปปิก

นายปุรณ์ ทองใส



วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาครุศาสตรมหาบัณฑิต
สาขาวิชาสถิติและสารสนเทศการศึกษา ภาควิชาวิจัยและจิตวิทยาการศึกษา
คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2565

| | |
|---|---|
| Thesis Title | AN ANALYSIS ON TRENDS OF RESEARCH TOPICS IN CIVIC EDUCATION USING DYNAMIC TOPIC MODEL |
| By | Mr. Poon Thongsai |
| Field of Study | Educational Statistics and Information |
| Thesis Advisor | Assistant Professor SIWACHOAT SRISUTTIYAKORN, Ph.D. |

Accepted by the FACULTY OF EDUCATION, Chulalongkorn University in Partial Fulfillment of the Requirement for the Master of Education

..... Dean of the FACULTY OF EDUCATION

(Associate Professor SIRIDEJ SUJIVA, Ph.D.)

THESIS COMMITTEE

..... Chairman

(Associate Professor Sungworn Ngudgratoke, Ph.D.)

..... Thesis Advisor

(Assistant Professor SIWACHOAT SRISUTTIYAKORN, Ph.D.)

..... Examiner

(Associate Professor KAMONWAN TANGDHANAKANOND, Ph.D.)

ปุรณ์ ทองใส : การวิเคราะห์แนวโน้มของประเด็นวิจัยทางพลเมืองศึกษาโดยใช้โมเดลไดนามิกทอปปิก. ( AN ANALYSIS ON TRENDS OF RESEARCH TOPICS IN CIVIC EDUCATION USING DYNAMIC TOPIC MODEL) อ.ที่ปรึกษาหลัก : ผศ. ดร.สิวะโชติ ศรีสุทธิยากร

วัตถุประสงค์ของวิทยานิพนธ์นี้คือการศึกษาการเปลี่ยนแปลงของหัวข้อวิจัยในด้านการศึกษาเพื่อสร้างความเป็นพลเมืองตั้งแต่ปี 2000 จนถึงปี 2020 และความสัมพันธ์ระหว่างหัวข้อวิจัยดังกล่าวกับปัจจัยด้านพื้นเพของตัวผู้วิจัยโดยนำเครื่องมือแบบจำลองสรุปหัวข้อบทความ Structural Topic Model ผู้วิจัยเก็บข้อมูลงานที่วิจัยที่เกี่ยวข้องกับการศึกษาเพื่อสร้างความเป็นพลเมือง รวมทั้งข้อมูลพื้นฐานของตัวผู้เขียนงานวิจัย จากฐานข้อมูลพื้นเพผู้วิจัย SCOPUS และฐานข้อมูลบทความ ERIC บทคัดย่อภาษาอังกฤษ ปีที่ตีพิมพ์ พื้นเพของผู้วิจัย และ h-index ของผู้วิจัย โดยใช้คียเวิร์ด “civic education” หรือ “citizenship education” หรือ “civics” ในการค้นหาบทความ บทความที่ค้นพบมีทั้งสิ้น 4917 บทความ หลังจากนำบทความมาเตรียมข้อมูลผ่านเทคทิคพื้นฐาน text mining ก่อนการวิเคราะห์ เหลือบทความที่เหมาะแก่การวิเคราะห์ทั้งสิ้น 4854 บทความ เครื่องมือแบบจำลองสรุปหัวข้อบทความ Structural Topic Model ถูกนำมาใช้ในการวิเคราะห์ข้อมูลบทคัดย่อ โดยเพิ่ม ปีที่ตีพิมพ์ และ พื้นเพของผู้วิจัย เป็นตัวแปรภายนอก เพื่อศึกษาหัวข้อวิจัยของบทความ ผลการวิจัยพบว่า 1) การค้นคว้าเรื่องการศึกษาเพื่อสร้างความเป็นพลเมืองที่เกี่ยวข้องกับการวิเคราะห์เชิงปริมาณ และ งานวิจัยที่เกี่ยวข้องกับมิติด้านการเมือง มีความสำคัญลดลงในรอบสองทศวรรษ ในขณะที่หัวข้อวิจัยที่เน้นไปทางการพัฒนาทักษะมีแนวโน้มจะได้รับความสนใจมากขึ้น 2) หัวข้อวิจัยเรื่องการศึกษาเพื่อสร้างพลเมืองโลกการศึกษาว่าด้วยสำนึกในชาติ และ งานวิจัยเชิงปริมาณได้รับความนิยมเป็นอย่างมากจากนักวิจัยทั่วโลก หัวข้อที่แนวโน้มจะได้รับความสนใจมากขึ้นจากการวิเคราะห์ข้อ 1 อย่าง การศึกษาโดยการบริการสังคม ได้รับความสนใจเป็นพิเศษจากนักวิจัยจากยุโรปตะวันออก การศึกษาโดยการบริการสังคมยังได้รับความสนใจมากขึ้นจากนักวิจัยตะวันตกตั้งแต่หลังปี 2015 เป็นต้นมา งานวิจัยในเชิงปรัชญาได้รับความสนใจเป็นพิเศษจากนักวิจัยในลาตินอเมริกาและเอเชีย แต่ไม่พบความแตกต่างเหมือนเปรียบเทียบกับบทความที่มีนักวิจัยหลายคนที่มีพื้นเพต่างกัน



| | | |
|---|---|---|
| สาขาวิชา | สถิติและสารสนเทศการศึกษา | ลายมือชื่อนิสิต ..... |
| ปีการศึกษา | 2565 | ลายมือชื่อ อ.ที่ปรึกษาหลัก ..... |

## # 6280092927 : MAJOR EDUCATIONAL STATISTICS AND INFORMATION

KEYWORD: Structural Topic Model, Text Mining, Civic Education

Poon Thongsai : AN ANALYSIS ON TRENDS OF RESEARCH TOPICS IN CIVIC EDUCATION USING DYNAMIC TOPIC MODEL. Advisor: Asst. Prof. SIWACHOAT SRISUTTIYAKORN, Ph.D.

The aim of this thesis is to study the trend of civic and citizenship education research from 2000 to 2020 and the influence the regional background of researches has on the research discussion. Relevant data is collected from ERIC and SCOPUS database. This includes abstracts, published year, regional background of researchers, and author h-index. The keywords used are “civic education” or “citizenship education” or “civics”. There are 4917 papers extracted in total. Upon doing further preparation, 4854 articles are prepared for analysis. We apply Structural Topic model (STM) technique to the abstracts with covariates including the published year and the continents the researchers are from. We found that 1) quantitative research in civic education and politically oriented topics are in decline whereas topics concerning skill development and service learning program are on the rise, 2) In terms of regional background global citizenship education, quantitative research, and national consciousness education are amongst the most popular topics across regions. The rising topic like service learning programs however particular is interested by researchers from Eastern Europe, and Western researchers from 2015 onwards. Philosophical discussion is mainly focused on by Latin American and Asian researchers but the difference are non-significant when compared to papers that are published by researchers from different regional background.



| | | |
|---|---|---|
| Field of Study: | Educational Statistics and Information | Student's Signature ..... |
| Academic Year: | 2022 | Advisor's Signature ..... |

# ACKNOWLEDGEMENTS

I am deeply indebted to my advisor, Ajarn Siwachoat Srisuttiyakorn, for helping the helpless in and encouraging me to reach beyond the scope of my ability. I would also like to show gratitude towards the two thesis committee members, Ajarn Sungworn Ngudgratoke and Ajarn Kamonwan Tangdhanakanond, and Ajarn Prapasiri Ratchaprapapornkul for insightful remarks and helpful suggestion. From Ajarns, I learned many valuable skills of how to conduct proper research and learn from others.

Last but not least, friends and family despite not directly involved with the grueling process of writing the dissertation did provide essential workaround: N' Jiramate for helping me with the iThesis system, P' Rum for providing useful information, my mother and father for words of encouragement, and many others. Without a community, a human being would find it hard flourish.

Poon Thongsai

# TABLE OF CONTENTS

Page

ABSTRACT (THAI) ..... iii

ABSTRACT (ENGLISH) ..... iv

ACKNOWLEDGEMENTS ..... v

TABLE OF CONTENTS ..... vi

List of Table ..... viii

List of Figures ..... ix

Chapter 1 Introduction ..... 11

1.1 Background ..... 11

1.2 Research Questions ..... 12

1.3 Research Objectives ..... 12

1.4 Scope of Research ..... 13

1.5 Definitions ..... 13

1.6 The Significance of the Research ..... 14

Chapter 2 Literature Review ..... 15

2.1 On the Concept and Practice of Civic Education Research ..... 15

2.2 Change in Research Methodology ..... 19

2.3 Bag-of-Words (Bow) ..... 21

2.4 Topic Modelling ..... 22

2.5 Structural Topic Model (STM) ..... 23

2.6 Examples of Topic Model Application in Education Research ..... 24

Chapter 3 Data Pre-processing and Proposed Model ..... 26

3.1 Data Extraction ..... 26
3.2 Text preprocessing and model implementation ..... 27
3.3 Model Evaluation ..... 28
Chapter 4 Model Result ..... 30
The following show the result after model application and explanation of certain results. ..... 30
4.1 Descriptive information ..... 30
4.2 Model selection ..... 32
4.3 Topics and Topic Content ..... 32
4.4 Topic Regional Differentials and Topic Trends ..... 50
Chapter 5 Conclusion and Discussion ..... 1
5.1 Summary of Research ..... 1
5.2 Discussion ..... 2
5.3 Further Suggestion ..... 3
REFERENCES ..... 4
VITA ..... 13

# List of Table

**Page**

Table 1: the description and the word components of each topic ..... 31

Table 2: the description and the word components of each topic ..... 31

Table 3: the description and the word components of each topic ..... 35

Table 4: Summary of topics interest based on authors' background ..... 54

Table 5: Summary of topics interest based on authors' background as compared to co-authors papers ..... 55

# List of Figures

**Page**

Figure 1 *the plate notation of LDA model. The word w is shades as it is the only observed variable whereas the rest are latent* ..... 22

*Figure 2 the plate notion of the improved model Structural Topic Model (STM)* ..... 24

*Figure 3 A snapshot of R code used to pre-process text data before analysis.* ..... 28

*Figure 4 A snapshot of R code used to run our candidate models for further selection. Note that we set "prevalence" to include factor the published year of the paper (year) and the regional backgrounds of the authors (region_dummy) in order to studt how these covariates affect the topics of discussion* ..... 28

Figure 5 the line curves showing the number of published papers by regional background of researchers ..... 30

Figure 6 the smooth curve and scatter plot showing the trade-off between semantic coherence and exclusivity of each model ..... 32

Figure 7 the bar chart of topic prevalence on average across our corpus ..... 33

*Figure 8 the subplots showing words and their probability in each topic* ..... 34

Figure 9 the correlation table showing the correlations amongst topics ..... 34

*Figure 10 Scatter plot showing average topic popularity in 2020 and the rate of change in topic popularity* ..... 50

Figure 11 the long run trend of each topic as generated from the model ..... 51

*Figure 12 the chart of decision tree to study the interaction between the published year and the regional background Moving to the left branch is for answer "yes" while going to the right branch is "no"* ..... 52

*Figure 13 the treemap showing average topic prevalence in each region. The larger the box, the more prevalent each topic is in the region* ..... 53

*Figure 14 the topic comparison across different regional background* ..... 53

*Figure 15 the topic comparison to the co-author group* ..... 54

# Chapter 1 Introduction

In this chapter, we begin by outlining the introduction and the scope of this research. Basic information about what the research is about is also provided.

## 1.1 Background

Globalization has a tremendous effect on how each country proceeds with the education of their citizens(Bakhtiari, 2011).In particular, the consequence for civic education means that more and more people are interdependent with people from different socio- economic and cultural background through the advent of electronic and telecommunication technology. This phenomenon lends itself to question the early concepts on civic education that are tied mostly to a particular nation-state and culture, and also the actual practices of civic education which are inflexible to change and unable to use the available resources properly (DeMaine, 2002).

Ideologically speaking, civic education in particular is a way for the state to legitimize a certain set of ideology and needs to the population (Frazer, 2008). Hence it is inevitable to see how civic education is a congruence of two different disciplines, which are politics and education, whose connection has not been well appreciated much (Samudavanija, 2557). This is quite apparent when one sees how 'political education' and 'political socialization' are used quite interchangeably to civic education. In terms of implementation, civic education requires continuous effort and commitment to replenish citizens' understanding of civic values. This implies the broad scope of civic education that goes beyond the confine of schools or universities.



This is why research on civic education is crucial in organizing and problematizing aspects in the implementation and the content of civic education in such society. Together with close studies of education policies and cultural socio-economical context, this allows us to understand the pros and cons of each element and the overall picture of our civic education, in hope to make it better suit for our fast changing society (Bray and Thomas, 1995). Furthermore, a comparative study on the nation state level also allows us to gain insight into how certain policies and frameworks are successful in a certain social context, and how that could be implemented in a different set up for better result (Kerr,2012).

Furthermore, not only is understanding the research crucial to development and improvement of civic education but also learning about the changing discussion of research topics. Over time, interests on specific topics may vary dependent of many factors. They could be the economic requirement of human capital, or the societal need to instill new values into the new generation. This allows us, from the perspective of researchers, to gain insights into the development of civic education enacted in different society. Hence, an in depth analysis on the trends of such 'hot' topics can elucidate the state of civic education and how each society enacts changes through education in a way to open up a possibility in further research.

Despite numerous studies on civic education (Kennedy,2012: Corgan and Morris, 2001: UNDP, 2014: Chow and Kennedy, 2015: Zhu, 2018), the overall longitudinal trend of such studies in comparison with different contexts of each country is still lacking. Civic education or citizenship education is an essential development of democratic society. In order for us to create such an eligible development process in an ever changing time, an overall understanding of the direction of the research in civic education is crucial. However, in the age of Big Data and ever increasing amount of available data, the traditional methodology may prove unsustainable

Topic modelling like the Latent Dirichlet Allocation or LDA has proved useful for researchers as it allows a large amount of text data to be analyzed with ease. Moreover, it manages to avoid subjective bias that may occur in the traditional expert-driven method. With an ever increasing amount of data for analysis and the nature of intertwining relationships between the topics discussed and other external subjective factors beyond the text, the advent of automated text analysis proves essential for researchers to understand the changing trends of discussion.



## 1.2 Research Questions

- How do the trends of research topics in civic education change over time?
- How do authors' background affect the discussion surrounding civic education?

## 1.3 Research Objectives

To explore the changing trend of research topics in civic education between 2000 and 2020 and to study the different interest in research based on authors' background

## 1.4 Scope of Research

We will be conducting a time-series comparative and quantitative analysis where our sample consists of the abstracts and introductions of research papers in the past 20 years from database ERIC and SCOPUS. Abstracts are selected because there are very easy to extract since they are in English. They are also ideal for comparison and computational analysis. All the concerning papers from 2000 to 2020 are extracted from both ERIC and SCOPUS. We will also collect all the background information of each individual article (e.g. year of publish, authors' names, h-index) and also that of the journals themselves. We follow Manning and Edwards (2014) on the essential keywords used to identify relevant articles on civic education, consisting words like "civic", "citizen", "politics" for instance.

We are interested in the variation of research topics over time and the evolution of word probability. A collection of words in a topic and how each word and topic changes over time allow us to learn about the nature of discourses surrounding the research of a particular field. This is done effectively previously in other works (Zheng, B., McLean, D.C. & Lu, X,2006: Lee, H., Kang, P,2018: Daenekindt and Huisman, 2020). The trends from each country is then compared for other insights by also using the composition of civic education by Carreto, Haste, and Bermudez (2016) and the concept of citizenship by Kanhe and Wertheimer (2004). More details are discussed in the next section.



## 1.5 Definitions

**Civic education** concerns the process led by the state through the educational system to foster generations of citizens with values and responsibilities deemed relevant to that society. The research on civic education comprises insights from a variety of research disciplines like psychology, education, political science, sociology and policy research etc.

**Topic Model** is a statistical time-series analyzing tool able to learn the latent topic discussion within a set of input text data. This allows researchers to gain insights into what important subjects are embedded in the text data across time with little effort and resources.

**Research topics or topics** here specifically mean a collection of words that are the essential components of the topic discussion. This basically follows the main assumption of topic modelling so that it can analyze words as a way to understand topics as a latent variable.

**Trends** in this research refer to the changing 'popularity' of a collection of words which associates with a set of topics. In other words, we look at how the likeliness of each word being part of a topic changes over time. This is done through topic modelling and lends us an opportunity to learn how each topic varies in a specific time period.

1.6 The Significance of the Research

The result should improve our understanding of the dynamic within the discourses of civic education research from the countries which have shown to excel in such field. This may allow researchers and anyone interested to gain another insight of one's society practice of civic education from comparing it to others, and hopefully led to a new direction of research or a new way to think about our relations in society

# Chapter 2 Literature Review

In this chapter we provide basic background of civic and citizenship education. We also outline the need for expansion of research methodology and the basic concepts of text analysis and machine learning methods

## 2.1 On the Concept and Practice of Civic Education Research

Here we will begin by discussing the theoretical background on the concepts of citizenship and civic education. This foregrounds the importance of civic education in each society. We also look at the previous research done on the international scale on the topic of comparative civic education. This allows us to see the limit of methodology in working with a large scale data and understand that a new computational tool is needed. Furthermore, the computational techniques of topic modelling and its relevance to our research are also briefly explored.

The concept of citizens and citizenship has been under scrutiny for centuries with no clear consensus among scholars, starting from Aristotle in Ancient Greece to the Enlightenment Age and the modern age. One can roughly distinguish two different models of citizenship: the republican and the liberal. Republican citizenship, whose lineage goes back to the Ancient Greece, emphasises the ability of self-rule of the public, and the interchageable capacity of ruling and being ruled. This means that republicans advocate the active participation in the political body as the utmost form of liberty,whether writing laws, becoming a representative, or deliberatng their political decisions. The liberal model of citizenship, on the other hand, is mostly interested in the legal aspect under the rule of law, that each citizen is all equally protected and has rights under the law. This model of citizenship has a long history especially during the Enlightenment Age, influencing the way political liberty is seen not as participartory but juridicial, as ‘freedom from interference from other individuals and the state”(J.S. Mill, 2002)

One can dissect further the partial elements of the concept of citizenship. Cohen (1999) suggests there are three interdependent aspects at play in the concept of citizenship. First is the juridical one which entails that all citizens are equally titled to rights and legal protection under the laws. Second is the political one which is interested in the equal deliberative participation of citizens in politics. And lastly, it is citizenship as a form of membership, a community tie which helps form citizen’s identity. Kahne and Westheimer (2004) also outlines three elements as well:

the justice-oriented citizen, the participatory citizen, and the personally responsible citizen. Although the definitions the first two aspects by Kahn and Westheimer are quite similar to Cohen, the last one, the personally responsible, is different from that as form of membership. The personally responsible citizen for them is defined as citizens who act responsibly within their community, like picking up litter or paying taxes.

Many studies have also problematized the very concept of citizens and both models of citizenship themselves. Scholars have questioned the underlined assumption of both the liberal and republican models, which is the separation between the public and the private. As in the republican model, according to Aristotle, the political sphere is seen as the epitome of liberty and equality, something that further completes human beings as the "political animals", while the private sphere is that of fixed nature and necessity, that there's no room for human deliberation. Many see this picture of politics exclude many people (slaves, women, and disabled people). They also argue that the dialectic relation between the private and the public/political: most of the "natural" misery within the private sphere, be it poverty, domestic abuse, or , can also be the consequences of the decisions made within the public sphere. And that many private individual issues, say, sexism and authoritarian family structure, must also be addressed on the social level and resolve with a collective action, which is something that the liberal model somewhat fails to conceptualize. Furthermore, A.Tyson and Choon Park (2008) inquires further on the relation between citizens and social justice, especially on the issue of race. Many focus on the effect of multiculturalism has on civic education as more and more people from different cultures are pouring in to Europe (Banks, 2008:Wong Fillmore, 1991: Appiah, 2006)



The teaching of civic education in Thailand varied greatly over the course of decades in terms of ideology and "Thainess" the state wanted to instill in Thai people (Mansap, 2550). Although Thailand had changed from absolute monarchy to constitutional democracy for almost 90 years now, a corpus of studies in citizen is still deemed somewhat alien to our society (Kalyanamitra, 2019). Many studies have also confirmed this. Eugenie Merieau et al (2011) finds that Thai society in general lacks the ability to think critically and voice their political opinion, and general understanding of how political facilities in the country function, all of which are essential elements in developing a democratic citizen in society. Furthermore, there seems to be a misunderstanding amongst the administrative wing of schools how they use schools as a way to foster democratic citizenship. Some schools are not allowed their students to participate in the

decision making process (e.g. how the rules should be made), and encourage authoritarian mindset instead of critical thinking (Boontinand, 2017:Tepnarin , 2019)

With respect to the theorization of civic education, Carreto, Haste, and Bermudez (2016) proposes four elements that are essential to civic education as follows:

- Civic understanding
- Civic skills
- Civic values, motivation and identity
- Civic action

Civic understanding includes general knowledge on the function of political body in one's society, social issues, and relationship amongst social groups. Civic skills entail all the abilities necessary to the participation of any social action (for instance, critical thinking, communication skills, or the ability to work with others). Civic values, motivation, and identity encompasses the development of values and identities within the process of deliberative participation in democratic society, most importantly the generational transfer of values and identity. Finally, civic action is the occasion in which political participation is possible, for example a consensus-seeking gathering within a community to try to solve certain issues.

The first two elements are crucially interdependent. Should we compare civic understanding to 'know-what', then civic skills are nothing but 'know-how'. This coincides with other researches which suggest the co-dependence between theory and practice in developing a democratic civic characters. For instance, Parker (2008) proposes two aspects which are 'democratic engagement' and 'political engagement' as two essential poles with which any civic education program needs to take into consideration seriously. As mentioned before that the nature of civic education is deeply connected with politics, the most important question one could ask in civic education is what sets of knowledge and ideology are being taught? And what sort of teaching programs are being used (Apple, 2018: Davies, 2008: Bannerji, 2011).

In terms comparative studies on civic education, they have gained interest over the years, as can be seen in numerous research comparing civic education in different countries. For example, Barber and Torney-Purta (2013) compares the ideology of nationalism and the support for

immigrants in the youth from 25 countries. Knowles (2015) explores the relation between “Asian values” and democratic citizenship in South Korea. Research in civic education is inter-disciplinary, as it requires insight from sociology, political science, psychology etc. One of the most beneficial byproducts of comparative studies is the awareness of different practices in civic education. This allows us to see the results of various implementations of civic education that could not be done in one country alone. Moreover, this gives us an opportunity to question our assumptions that we may have on our own practice of civic education that are left untouched before international comparisons(Hahn-Alvier-Martin, 2008).

As comparative studies on civic education grow, the need for large-scale data is inevitable. Two of the most comprehensive sets of such data are by the International Association for the Evaluation of Educational Achivement or IEA. These are the Civic Education Study of 1999 (CIVED:99) and The International Civics and Citizenship Education Study of 2009(ICCS:09) (Barber,Knowles, and Torney-Purta,2018), where both data collect from surveys of students aged 14-14 from over 40 countries. They also include basic information on the teachers and the schools as well.

There are a few of studies which give an overall picture as to how these two data sets are being used. Di Stefano and Knowles (2015) lists all of the numerical studies on civic education which use the IEA data sets. Such studies are identified through Google Scholar engine and the database of ERIC: Institute of Education Service for additional studies before being individually summarised. Barber, Knowles and Torney-Purta(2018) proceeds in a similar fashion and classifies all such studies into four groups based on their goals of study. These four groups are open classroom climates, teaching and learning approaches, student identity, and profiles of citizenship norms and attitudes. Finally, Hahn and Alvier-Martin (2008) studies the trend of such studies specifically about political socialization.

Apart from research on the IEA data sets, there are also other scholar works comparing research topics on civic education. Kerr(1999) summarizes all the studies on the civic education concerning the curriculum. Bennion and Laughlin(2018) explores topics discussed under the umbrella of civic education in research papers in the Journal of Political Science Education. They want to study the changing trend of discussion over the first twelve years of the establishment of the journal. Geboers et al(2013) analyses 28 studies which involve the topic of the effectiveness of civic education on students in school. Castro and Knowles(2017) synthesizes the latent direction of research under the context of social, cultural and economic change by focusing on three

journals: Research in Social Studies Education, Education for Citizenship and Democracy, and Research in Civic Engagement in Youth

There are in particular many such studies done by researchers outside of Thailand. For instance, Kennedy (2012) compares the relation between civic values and civic understanding of the students in Hong Kong and Thailand. Corgan and Morris(2001) analyses the civic education programs in six countries including Thailand while taking into consideration education policies, the effect of such policies on the curriculums, and the actual implementation of those curriculums. Chow and Kennedy (2015) studies the 'future civic engagement' amongst Asian countries(Hong Kong, Indonesia, South Korea, Taiwan, Thailand). In UNDP (2014) we have a full detailed study of youth civic participation amongst 12 Asian countries. Lastly, Kalyanamitra (2019) provides a case study of civic education practices in German and Thailand.

2.2 Change in Research Methodology

Prior to the advent of automated text analysis like topic modelling, research methodology, as we have discussed, usually consists of case studies of relatively small sample. The corpus is then read through by a number of experts with a particular established framework in mind in order to decipher insights from the selected texts. There are many examples of such research over time (Chiva-Bartoll et al, 2019: Finkel et all, 2022: Fitzgerald et al, 2022). Despite incredibly useful, this traditional method has proven lacking in the face of ever increasing amount of available data for analysis as it also imposes tremendous cost in terms of time and personnel to researchers (Lindtstedt, 2019). Topic modelling is one such alternative as it allows a large amount of text data to be analysed with ease. (Blei et al, 2003).The application of topic modelling to social science study especially education is quite common (Manning and Edwards, 2014: Daenekindt and Huisman,2020). In order to understand the working of topic model, one must first inquire into the back process of natural language process particularly called text mining.

2.3 Text Mining

Text mining concerns the process of extracting insightful information from text data. This is done by allowing computers to automatically learn the latent patterns within the corpus of input text data and present us with useful information. Since most of the online data stored today are unstructured texts, this method is incredibly useful in learning and discovering new knowledge from the data with much less effort. It is incorporated extensively in the business world and the

academic world alike. It draws from many disciplines such as ‘information retrieval, machine learning, statistics, computation linguistics and especially data mining’ (Hotho et al, 2005).

More precisely, one can dissect text mining into two different phrases: ‘text refining’ and ‘knowledge distillation’(Chudambaram and Sumathy, 2013).Text refining refers to the extraction of unstructured text data to a selected form, meanwhile the next step is knowledge distillation which is where we decipher latent patterns within the data to gain new insights. The form derived from the text refining phrase can take many types. It can be document based or concept based. This then leads to many kinds on operations one can process to the inpiut data: the document based may give the opportunity to visualise or categorise, whereas the concept based may allow us to produce a predictive model which can be used for other set of texts in the future(Chudambaram and Sumathy, 2013).

However, as most of text data we’re interested in are unstructured, a pre-processing tool is required to cleanse it first before an analysis can be done. Natural Langauge Processing or NLP is one such tool which automatically tries to fully represent human languages in a way computers could understand by making use of the grammatical structure of language both syntax and semantics(Gupta and Lehal, 2009:Besancon and Rajman,1997). According to Nadkarni et al (2011), some of the common tasks in NLP include

- *Sentence boundary detection* is spotting out stopwords,abbreviations and titles. For examples, words like “if”, “the”, “and”, “Mr” ,”DIY” or “KAIST”. These words are basically the ‘noises’ in our input data which can affect the quality of our analysis if left untouched.

- *Tokenisation* is dividing out the text into distinct units,usually words or sentences. For example, we may cut up the sentence “I ate rice yesterday” into four distinct word units like “I”, “ate”, “rice” and “yesterday” for further analysis. This is quite essential if our research is to learn about the relation of words in the corpus. We can also delete punctutations here in this process.

- *Part-of-speech assignment* or *POS tagging* is marking each word to its part of speech. There are 8 different taggings : Nouns, Verbs, Adjectives, Adverbs, Preposition, Conjunction, and Interjection.

- *Decapitalising* is to decapitalise words, like “Good” is changed to “good”. This is also

important since the model we use in our analysis may not be able to understand the similarlity of the two.

- *Stemming* is the process of removing the affixes and suffixes from the words so that we get the root of the word. For instance, stemming may change “quicker” to “quick”, “hopeless” to “hope”, or “microbiology” to “biology”.Lemmatising is similar in concept to stemming. Lemmatisation allows us to return the form of word into their base root. For example, “ate” is lemmatised back to “eat”. “Went”,”goes”, and “Gone” are reduced back to “go”.

- *Chunking* is extracting structural phrases from the text by learning the POS tagging of each word in a sentence.

It is apparent that text mining allows us to gain further insights into a collection of large texts with less effort and resources compared to other classical methods. One of the particular techniques is to map out different discussions within a certain set of texts, which basically makes it possible to understand the topics mentioned within it.

2.3 Bag-of-Words (Bow)

Another essential step for text analysis is to transform words in corpus into vectors for further analysis. This approach is called Bag-of-Words or BOW for short. The idea is that we try to compartmentalize word occurrences based on our corpus so that our probabilistic analysis is possible. This means that we completely ignore grammar and word order. In other words, if our entire corpus consists of $d$ words in total. Our text, in our case a research paper, is transformed into a vector of dimension $d$ where each coordinate represents the number of times the associated words appear in that document. For example, suppose that the “dictionary” of our entire corpus is $['apple','papaya','banana','orange']$ and suppose our text consists of the word apple 4 times and the word banana 10 times. The vector representing this specific text is then $[4, 0, 10, 0]$ showing the frequencies of each word in the corpus respectively.

## 2.4 Topic Modelling

Topic modelling is a powerful statistical tool which makes possible the extraction of latent "topics" that present in a large corpus of text data. The 'topic' here is defined as a collection of words that are very likely to occur together across texts, where topic modelling maps similar patterns of word use and connects different documents with similar structure. Topic is hence assumed to be a probability distribution over words.

The main assumption of LDA is that each document represents a mixture, with a certain distribution, of topics which themselves consist of several distributed words. The observed words in documents are treated as results of a hidden probabilistic generating process which can be reversed-engineered by the model. In a more mathematical form as shown in Figure 1, we begin by drawing $k$ topics $\beta_1, \beta_2, \dots, \beta_k$ from the topic distribution modelled by Dirichlet distribution with parameter$\eta$. For each document $d$ , $\theta_d$ the respective topic-per-document distribution is also drawn from the Dirichlet distribution with parameter $\alpha$. Next, for each word $w$ in document $d$ we assign its topic called $Z_{d,n}$ drawn from multinomial distribution $Multinorm(\theta_d)$, and the observed word itself is drawn from $Multinorm(Z_{d,n})$ (Blei et al 2003).

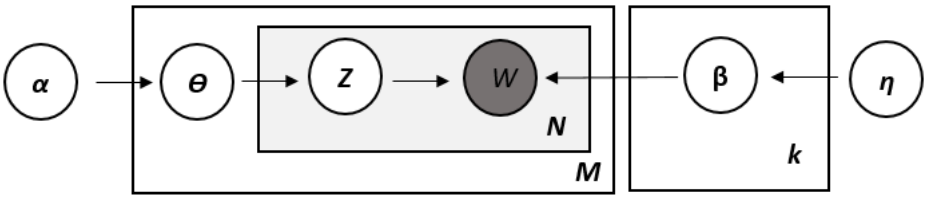


Figure 1 *the plate notation of LDA model. The word w is shades as it is the only observed variable whereas the rest are latent*

Although the LDA model allows us to gain incredible insights into the text data, there are some shortcomings. Firstly, the topics are themselves unrelated within the documents. This proves rather unrealistic, when some topics are usually discussed together. Second, the distribution of words within a topic remains constant. This is again rather idealistic when it comes to topics which

may have different approaches to discussing it, and therefore different set of associated words. Lastly, it only takes into account the text data alone but no other contextual data which may prove influential in determining the latent topics.

## 2.5 Structural Topic Model (STM)

The Structural Topic Model or STM is an improvement on the traditional topic model (i.e. the Latent Dirichlet Allocation or LDA) that allows the inclusion of covariate information of the documents or "metadata" into topic analysis (Roberts et al, 2019). Similar to LDA, a document is assumed to be a distribution of $K$ topics. Whereas in LDA the topic proportions of each document ($\theta$) are fixed and set in advance from the Dirichlet distributions, in STM $\theta$ can be correlated, and the prevalence of these topics is derived under a generalized linear model with covariates $X$ from the metadata. In other words, we can estimate $\theta$ under the assumption $\theta \sim LogisticNormal(\gamma X, \Sigma)$ where $X$ represents the exogenous covariates mentioned and $\gamma$ the weights, sampled from a Half-Cauchy distribution, which normalize the exogenous covariates to resemble topic prevalence. Next, we can generate the topic for each word based on a multinomial consisted of these topic proportions. That is, $z_{d,n} \sim Multinomial(\theta_d)$ for every $n$-th word in document $d$ where $\theta_d$ represents the topic prevalence of document $d$ estimated previously. Conditional on this topic $z_{d,n}$, a word is selected from a multinomial distribution of words parametrized by $\beta$ derived from the logit deviation from the baseline word frequencies $m$. That is, $\beta \propto \exp(m + \kappa)$ where $\kappa$ stands for the relationship between topics and covariates (Roberts et al, 2014)



In sum, there are at least three advantages in using STM compared to the traditional model: (1) Topics can be correlated, (2) Topic content, or the distribution of words within a topic is varied, and (3) Topic prevalence is also modelled with external data beyond the texts. These benefits are useful especially in social science research, where discussion of the same topics can be of a different perspective, hence consisted of different set of words. Furthermore, the ability of the model to include exogenous variables beyond the input text to analyze topics is also interesting as topics are much more likely to be influenced by other factors beyond the text itself.

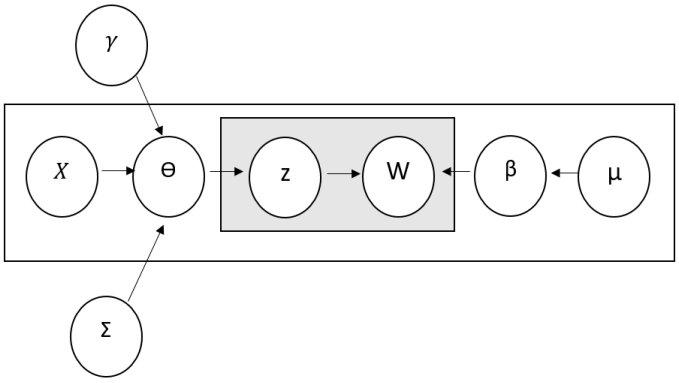


*Figure 2 the plate notion of the improved model Structural Topic Model (STM)*

2.6 Examples of Topic Model Application in Education Research

Prior to the advent of automated text analysis like topic modelling, research methodology, as we have discussed, usually consists of case studies of relatively small sample. The corpus is then read through by a number of experts with a particular established framework in mind in order to decipher insights from the selected texts. There are many examples of such research over time (Chiva-Bartoll et al, 2019: Finkel et all, 2022: Fitzgerald et al, 2022). Despite incredibly useful, this traditional method has proven lacking in the face of ever increasing amount of available data for analysis as it also imposes tremendous cost in terms of time and personnel to researchers (Lindtstedt, 2019). Topic modelling is one such alternative as it allows a large amount of text data to be analysed with ease. (Blei et al, 2003).The application of topic modelling to social science study especially education is quite common (Manning and Edwards, 2014: Daenekindt and Huisman,2020).



Fitzgerald et al (2021) studies the current nature of civic education research concerning youth participation in the United States. The main concern of the paper situates itself within the shifting political climate in the US, and the need to have an improved understanding and a more efficient way to measure civic education in order to foster civic engagement in youth. The sample articles are searched electronically on a set of keywords, and then inspected individually by reading the abstracts and the introduction. Most importantly, all the research papers in this study are grouped into three different aspects: the methodology, the themes related to youth participation, and the level of education studied. In each group, sub-groups are divided further based on a grounded theory approach.

Quite similarly, Manning and Edwards (2014) examine the current state of civic education concerning youth 'normative' political participation like voting, being part of a political party, or signing a petition. This is achieved through a traditional means of screening manually through relevant quantitative research articles. Data is checked and insights are extracted by researchers. Next, Daenekindt and Huisman (2020) study multifaceted research in higher education. They apply the topic model on the corpus of abstracts from a collection of selected journals. They suggest further that their methodology is to refine from previous systematic reviews which either fall short due to the lack of large scale in-depth content analysis.

Nevertheless, LDA itself has some caveats which led to various improved models over time. Structural Topic Model or STM is one such alteration which seeks to improve the LDA by including exogenous factors and varied topic content and topic relations into the analysis. For instance, Morales and Paz (2022) analyses the text arguments of political debates during the Chile's constitutional process in 2015 in order to study what drives civic participation. Together with the written texts, socio-economic index of citizens' regional backgrounds are also included into the model. Mi et al (2020) applies STM to articles in order to understand the trends and latent discussion within research on students' conceptual understanding in science education. Shi et al (2022) uses STM to study the discourses concerning sex education on an online website by paying close attention to external factor influence upon such discussion like gender difference, expected educator and expected receiver of such education. Although there does seem to have a number of social science research which incorporates the STM technique into their analysis, research in civic education is yet lacking, which opens room for further analysis.

# Chapter 3 Data Pre-processing and Proposed Model

In this chapter we outline the process of data extraction, data preparation, and model application.

## 3.1 Data Extraction

Articles with relevant keywords are extracted from the ERIC and the SCOPUS databases. By utilizing the native search syntax 'subject' available at the ERIC database, we can extract all the documents whose keywords contain "civic education" or "citizenship education" or "civics". Hence, the search string is as follows: subject: "civic education" OR subject: "citizenship education" OR subject: "civics". Once obtained, the data outside of the years 2000 and 2020 is discarded. Similarly for SCOPUS, we insert the following search syntax: KEY ( "Civic Education" ) OR KEY ( "Citizenship Education" ) OR KEY ( "Civics" ) ) AND PUBYEAR > 1999 AND PUBYEAR < 2021. Authors' names, abstracts, titles of the papers, article keywords, names of the published journal, publication types, published years, and the published years are called. Articles whose abstracts are less than 30 words are removed.

Next, we apply the program available in the python package *pybliometrics* to extract the further information on the authors. It allows us to scorch the SCOPUS database of all the associated data with the authors like their fields of interest, affiliations and so forth (Rose and Kitchin, 2019). Firstly, using the title of the articles, the authors ID are matched and then used to acquire their affiliations and fields of interests. Second, the data left undecided from the previous session is now accounted for properly by using the author's names. However, there may be still be some unresolved cells from the dataset due to similar names or a lack of available information in the database. The third step is to check each missing row individually. This consists of the following steps:

1. Look up the article name and obtain the actual paper file. The author's affiliations are often included in the paper itself. We then use this information to look for the desired affiliations using the python program.
2. If the paper itself is inaccessible or does not have any information of the authors, the authors' name themselves are searched up in aforementioned websites. If the authors' CV or list of previous publication exist, we can examine whether it contains the articles we are looking for. If

so, we use their association which are most likely included in their CVs to insert in the python program.

3. If all above fail, we left the cells as none.

Lastly, we categorize the author affiliations by regions. This is based on the way the SCIMAGO Journal ranking website classifies their journal places of origins. It consists of 8 regions in total: 'Latin America', 'Middle East', 'Northern America', 'Africa', 'Eastern Europe', 'Asiatic Region', 'Western Europe', 'Pacific Region'. For our further analysis however, we additionally group certain regions together for easier interpretation and less sparse data. That is, we group Northern America and Western Europe together as 'The West', whereas Asiatic Region and Pacific Region are combined as 'Asia'. We also flag any document which has at least two different regional researchers as "co-author". We may use "co-authors" as a benchmark to study how researchers from each region deviate.

3.2 Text preprocessing and model implementation

After finishing garner all the needed information, the text preprocessing is the final step before executing topic model. We apply the standard method of text cleansing in natural language processing available in the python package *gensim* to the abstracts and the titles. As shown in Figure 3 below, we apply the stemming process, remove unnecessary input like punctuation, numbers, stop words, additional words like "civic" or "education", and short words whose length is under 5 to avoid noise in our model. Moreover, we also delete any words which may appear in more than 1000 documents to again avoid redundancy in our model (this is done by setting *upper.thresh* argument to 1000 as shown below). Having prepared the dataset, we make use of the *stm* package from R for our analysis. This package is basically the scripted program which allows us to estimate Structural Topic Model (STM) and visualise the results. In terms of model specification, we use spectral initialization since it yields consistent result (Roberts et al, 2019). To run candidate models, this is done easily through a function called *searchK in* the R code as shown in Figure 4. It will run our candidate models and evaluate important metrics for model selection like semantic coherence and exclusivity which will be discussed shortly. Since our research is interested in how authors' background affects the discussion in each topic, we make use of STM model to include the regional affiliation of the authors as a covariate. Furthermore, since we are also interested in the trend of topic discussion over time we also include the variable published year into the model as well.

```
#Text Preprocessing
set.seed(420)
processed <- textProcessor(df$abstract1,onlycharacter=TRUE,ucp=TRUE,metadata = df,stem=TRUE,wordLengths = c(5,Inf),striphtml=T,
                           customstopwords = c('abstract','civic','education','citizenship','quot','amp','apos','citizen','research','study'))
out <- prepDocuments(processed$documents, processed$vocab, processed$meta,verbose=TRUE,lower.thresh = 10,upper.thresh = 1000)
docs <- out$documents
vocab <- out$vocab
meta <- out$meta
```

*Figure 3 A snapshot of R code used to pre-process text data before analysis.*

```
cl<-makeCluster(detectCores()-1,type='SOCK')
registerDoSNOW(cl)
tic()
stm.search <-searchK(documents = docs,
                     vocab =vocab,
                     K=seq(5,50),
                     seed=420,
                     prevalence = ~s(year)+region_dummy,
                     data=meta,
                     init.type = 'Spectral',
                     verbose=T)
toc()
registerDoSEQ()
```

*Figure 4 A snapshot of R code used to run our candidate models for further selection. Note that we set "prevalence" to include factor the published year of the paper (year) and the regional backgrounds of the authors (region_dummy) in order to studt how these covariates affect the topics of discussion*



3.3 Model Evaluation

The next process is to evaluate our result from the model. Exclusivity is the indication of how likely the top words of a topic are also the top words of other topics. This means that a topic is exclusive if the most likely words associated with this topic do not appear as top words in any other topic. Semantic coherence on the other hand measures how cohesive the topic is by looking at how top words occur together. That implies that a topic is "coherent" if the highly probable words of that topic most likely appear together in documents. Semantic coherence correlates well with human judgment, however a high number of semantic coherence is easily achieved if there are only a few topics with common words (Mimno et al, 2011). For any topic there are the most $N$ probable words $V = (v_1, v_2, \dots, v_N)$. Semantic coherence can be calculated as $\sum_{i \neq j} \log\left(\frac{D(v_i,v_j)+1}{D(v_j)}\right)$ where the terms go through every possible combination of the distinct pair$(i, j)$, while $D(v_i, v_j)$ and $D(v_j)$ denote how often the word pair appear together in a

document and the frequency of word $v_j$ in a document, respectively (Mimno et al, 2011). Exclusivity, on the other hand, is calculated by the ratio of a word topic-specific frequency relative to the sum of such word frequency of other comparative topics (Bischof and Airoldi, 2012).

Once we have selected a model, a number of results will be of interest. First, correlation amongst topics (calculated from topic prevalence) is checked in order to understand the likelihood of topics being discussed together. Second, we can look at the topic prevalence across corpus and the composition of each topic by the probability of words in it. Third, we analyze further the top words of each topic and the relation between topic words and our covariates.

Fourth, we study how our covariates which in this case are the time of publish and regional affiliations have any differential effect on topic of discussion. This is done in two folds. Regression analysis is applied to the annual average topic proportion with the number of years before 2020 as the independent variable. This is to dissect the trends and the change in popularity of each topic up until 2020. Next, a decision tree model is also applied to understand the interaction between the two covariates in determining the topics of discussion. In the second part of the analysis, we select the topic with the maximum number of topic proportion to represent that particular document. This max topic factor will be used as our dependent variable in our decision tree model development. We split our corpus into an 80% training set and a 20% test set. This consists of 3889 articles for the training set and 965 for the test set. The training set will be used for decision tree model development. Repeated cross validation method is used with 5 folds repeated 3 times. The metric of accuracy will be used to gauge the model. This procedure is easily achieved using R package "caret" (Kuhn, 2008).



Lastly, we can study the preference of researchers from different regions and how it is compared to each other. We will also compare them to the co-author group, papers whose authors consist of multiple researchers from different background, to gain insight into how each region emphasizes on.

# Chapter 4 Model Result

The following show the result after model application and explanation of certain results.

## 4.1 Descriptive information

Figure 5 illustrates the trend of published articles across two decades based by the regional affiliations of the authors .Initially there are in total 5132 articles from the ERIC database. Articles whose abstracts are less than 30 words are removed, as a result only 4854 articles remaining as our input data. The following bar chart illustrates how the articles are spread out over two decades. Moreover, there are 3760 articles after 2010 whereas only 1372 in the decade before. This trend may suggest the growth of research in civic and citizenship education over the years. However, this may also indicate simply the improvement of data management of the databases. In terms of affiliations, 75.09% of the entire documents have authors whose affiliation is based in Northern America or Western Europe while the remaining 1245 articles belong to authors outside of the West. Note also that some articles may have authors from different regions. The internationally collaborated papers (i.e. the author affiliations includes both from the West and outside) only account for 6.2% of the Western-affiliated documents.

Furthermore, when we look at the average author h-index of each paper subject to the authors' regional background. Note that we use average for author h-index since some papers have more than one author. The top keywords based on the regional background are also shown to provide. Both do not seem to suggest any particular difference amongst different regions. These are provided in Table 1 and Table 2 below.



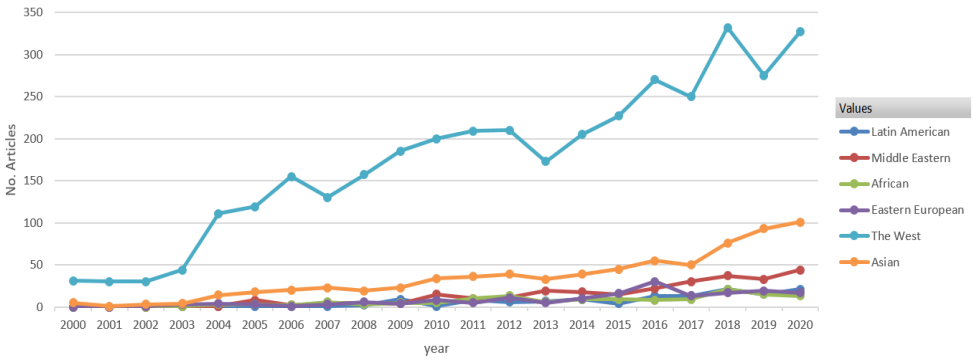


*Figure 5 the line curves showing the number of published papers by regional background of researchers*

*Table 1: the description and the word components of each topic*

| Author h-index by regional background | mean | sd | min | max |
|---|---|---|---|---|
| Africa | 1 | 0 | 1 | 1 |
| Asia | 1 | 0 | 1 | 1 |
| Eastern Europe | 1 | 0 | 1 | 1 |
| Latin America | 1 | 0 | 1 | 1 |
| Middle East | 1 | 0 | 1 | 1 |
| The West | 1 | 0 | 1 | 0 |
| co-author | 2.04 | 0.19 | 2 | 3 |

*Table 2: the description and the word components of each topic*

| | Top 10 words arranged from highest frequencies |
|---|---|
| Africa | education,foreign,school,citizenship,social education,civics,socialcountries citizenship, foreign education, democratic education |
| Asia | education,citizenship,foreign,school,civic,foreign education,global,citizenship education, countries citizenship |
| Eastern Europe | education,civic,citizenship,foreign,school,participation,political,secondary,eeducational |
| Latin America | education,citizenship,civic,language,school,social,international,foreign,second,educational |
| Middle East | education,school,citizenship,foreign,social,grade,teacher,curriculum,countries,citizenship countries |
| The West | education,school,citizenship,civic,foreign,social,citizenship education, social education, secondary |
| co-authors | education,citizenship,school,foreign,civic,foreign education,language,global,countries citizenship, cultural |

## 4.2 Model selection

Following Roberts et al (2014), we produce the plot between the exclusivity against the semantic coherence for each model is presented in Figure 6 below. The ideal topic number should be at the “knee” of the curve, that neither semantic coherence nor exclusivity dominates, as this suggest an optimal trade off option. It appears that a cluster of topic number between 10 and 19 situates around such frontier of the curve. Moreover as can be seen from the curve, between topic 15 and topic 17 there is substantial decrease in semantic coherence (a leftward horizontal distance) but insignificant increase in exclusivity (an upward vertical distance). That is, by sacrificing semantic coherence the marginal change in topic exclusivity is negligible when the number of topics increased from 15 to 17. Note also that the semantic coherence and exclusivity for both topic 15 and topic 16 are almost identical. Therefore, for the sake of parsimony, we will choose our number of topics 15. Note that there is no objectively correct number of topics, this model allows one to delve into the structure of the texts through topic adjustment.

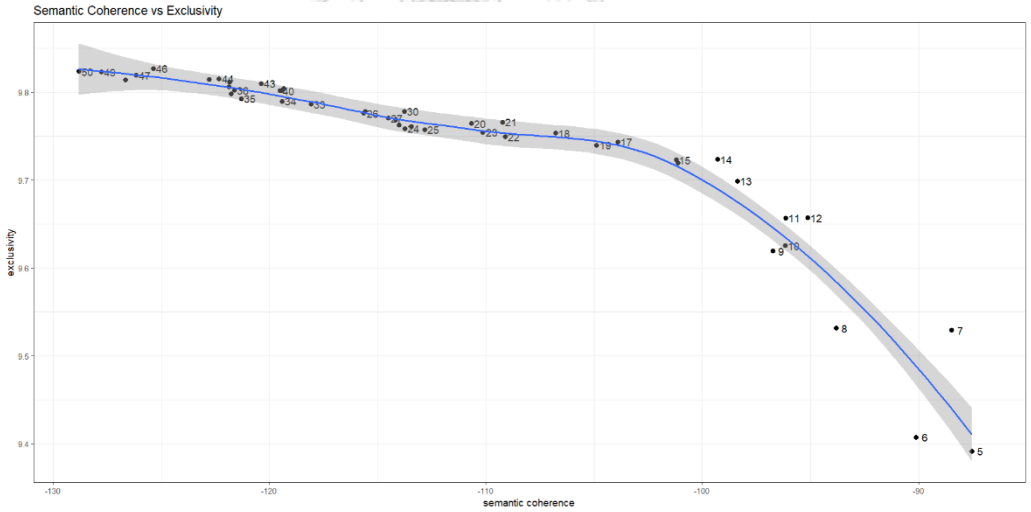


*Figure 6 the smooth curve and scatter plot showing the trade-off between semantic coherence and exclusivity of each model*

## 4.3 Topics and Topic Content

To introduce the resulted topics, the topic prevalence is provided below how in general each topic appears in discussion. This is basically the average topic proportion across corpus and between 2000 and 2020. The topic words by highest probability are also listed to further understand the composition of each topic. This is again to understand the composi These are provided in Figure 7 and Figure 8 below. We will further delve deeper into the word composition of topics later on.

We can further investigate the most significant words associated with each topic. Beyond the traditional topic modelling method of ranking them by their probabilities, there are a number of other criterions at play. FREX or Frequency-Exclusivity is metric that emphasizes words that appear frequent and also exclusive to a particular topic. This is because in order to discuss any topic certain words may crop up time and time again, yet it is not exclusive to a particular topic (Bischof and Airoldi, 2012). For example, when discussing education, we would expect words like "student", "education", or "teach" to appear rather often in any topic we model. This is depicted for all topics in Table 3.

Since we allow our covariate (authors' background) to have an effect on the words in each topic, we provide lists of words which are the result of topic covariate interaction as in STM model. Note that the words shown below may appear incorrect due to the stemming process in data preparation process. Furthermore, we can study the correlation amongst the topics themselves which are shown below in Figure 9. Positive correlations between topics indicate that both topics are likely to be discussed within a document. Since we are only interested in only positive correlations amongst topics, we will ignore any negative correlation and replace it with zero.

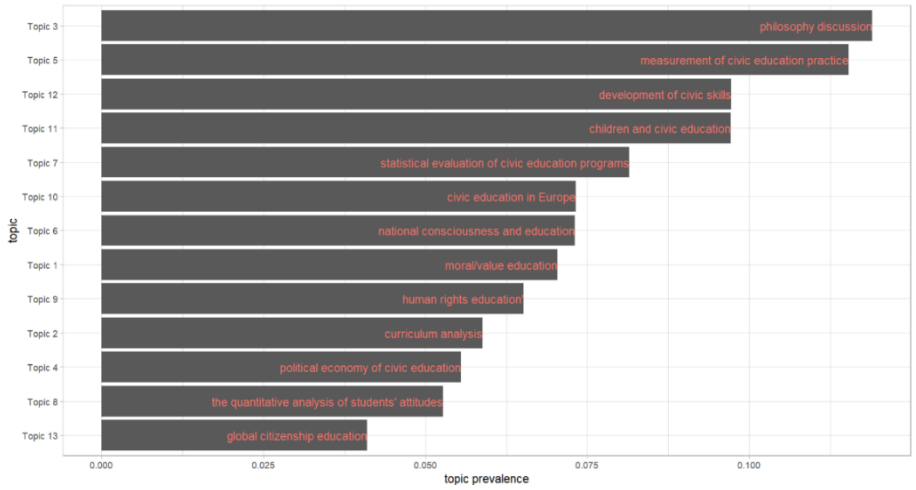


*Figure 7 the bar chart of topic prevalence on average across our corpus*

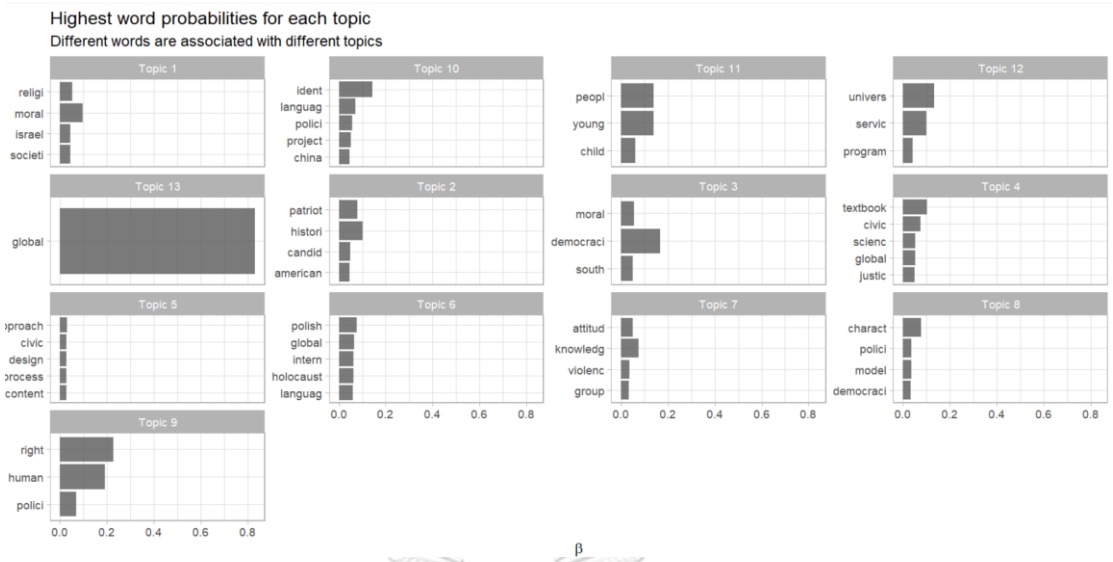


*Figure 8 the subplots showing words and their probability in each topic*

| | 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 1 | 1 | | | | | | | | | | | | | | |
| Topic 2 | 0 | 1 | | | | | | | | | | | | | |
| Topic 3 | 0.031152 | 0 | 1 | | | | | | | | | | | | |
| Topic 4 | 0 | 0 | 0 | 1 | | | | | | | | | | | |
| Topic 5 | 0 | 0 | 0 | 0 | 1 | | | | | | | | | | |
| Topic 6 | 0 | 0 | 0 | 0 | 0 | 1 | | | | | | | | | |
| Topic 7 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | | | | | | | | |
| Topic 8 | 0 | 0 | 0 | 0 | 0 | 0 | 0.068678 | 1 | | | | | | | |
| Topic 9 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | | | | | | |
| Topic 10 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | | | | | |
| topic 11 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | | | | |
| Topic 12 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.01222 | 1 | | | |
| Topic 13 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | | |
| Topic 14 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.039803 | 0 | 1 | |
| Topic 15 | 0 | 0 | 0 | 0 | 0 | 0.08168 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |



*Figure 9 the correlation table showing the correlations amongst topics*

*Table 3: the description and the word components of each topic*

| No. | Description | FREX words | Topic-covariate Interaction |
|---|---|---|---|
| 1 | moral/value education | christian, moral, philosoph, multicultur, spiritu, relativ, inculc | Group Africa: dissent, empathi, conscienc, incid, phase, program, didact<br><br>Group Asia: indigen, deliber, mathemat, profession, northern, interrog, socialis<br><br>Group co-author: apartheid, muslim, cosmopolitan, confession, ubuntu, faith, authent<br><br>Group Eastern_Europe: movement, ident, phenomenon, overcom, globalis, class, section<br><br>Group Latin_America: state, regress, percentag, attain, senior, thirti, select<br><br>Group Middle_East: decoloni, constrain, matrix, interfac, manag, linguist, languag<br><br>Group west: ecolog, blend, matrix, onlin, state, enrol, strategi |

| | | | |
|---|---|---|---|
| 2 | curriculum/content analysis | patriot, russia, soviet, ritual, russian, pledg, commemor | Group Africa: percentag, percept, attitud, great, presid, questionnair, democraci<br><br>Group Asia: operationalis, crime, critic, cultiv, colleg, enhanc, lectur<br><br>Group co-author: virtual, deviat, sexual, ident, entir, undergradu, youth<br><br>Group Eastern_Europe: renew, conscious, enquiri, metaphor, volunt, suprem, specialist<br><br>Group Latin_America: modal, psycholog, futur, predict, charg, femal, themat<br><br>Group Middle_East: ration, crisi, english, philosophi, literaci, modul, joint<br><br>Group west: modul, islam, literari, ration, democratis, secular, joint |

| | | | |
|---|---|---|---|
| 3 | philosophical discussion | liber, philosophi, truth, departur, plural, ubuntu, mouff | Group Africa: learner<br><br>Group Asia: singapor, asian, confucian, coach, arendt, israel, autonomi<br><br>Group co-author: japan, japanes, modul, resist, zealand, impos, histori<br><br>Group Eastern_Europe: korea, korean, holocaust, muslim, neutral, share, lifelong<br><br>Group Latin_America: mobil, tertiari, scholarship, typolog, embrac, zealand, liber<br><br>Group Middle_East: junior, simul, score, indonesia, digit, disposit, intellig<br><br>Group west: taiwan, taiwanes, music, mainland, island, migrat, daili |

| | | | |
|---|---|---|---|
| 4 | the political economy of civic education | econom, socio, neoliber, discern, market, cater, competit | Group Africa: elementari, partner, pakistan, prosoci, nativ, latent, event<br><br>Group Asia: indonesian, quasi, thailand, religi, indonesia, style, wisdom<br><br>Group co-author: consensus, villag, activist, percentag, consult, physic, indonesia<br><br>Group Eastern_Europe: resili, china, mainland, financi, identif, unemploy, spatial<br><br>Group Latin_America: australian, comput, normat, internet, migrant, spatial, environment<br><br>Group Middle_East: canadian, peacebuild, lifelong, civil, immigr, unpack, settlement<br><br>Group west: circl, multilevel, concentr, competit, cosmopolitan, politi, border |

| | | | |
|---|---|---|---|
| 5 | measurement of civic education practice | video, visual, method, triangul, compris, design, qualit | Group Africa: sport, justif, formal, undergradu, classroom, showcas, academ<br><br>Group Asia: intercultur, novel, session, visual, english, christian, literaci<br><br>Group co-author: taiwan, volunt, realist, financi, charact, english, japan<br><br>Group Eastern_Europe: swedish, modul, ration, agonist, regular, orthodox, younger<br><br>Group Latin_America: holocaust, religion, christian, african, tomorrow, ubuntu, australian<br><br>Group Middle_East: argentina, indigen, mathemat, palestinian, canadian, israel, programm<br><br>Group west: profession, specialist, scholarship, target, partner, financ, formal |

| | | | |
|---|---|---|---|
| 6 | national consciousness and education | postcoloni, china, chines, mexican, stereotyp, brazil, perpetu | Group Africa: intercultur, british, synthesi, religion, nativ, japanes, heart<br><br>Group Asia: passag, korean, crime, child, wrong, assess, slovakia<br><br>Group co-author: track, socioeconom, volunt, legal, singapor, variabl, colombia<br><br>Group Eastern_Europe: deliber, simul, engin, voluntari, academi, secular, track<br><br>Group Latin_America: islam, muslim, pakistan, dutch, turkish, repertoir, turkey<br><br>Group Middle_East: woman, discours, network, incid, ecolog, connect, efficaci<br><br>Group west: academ, dissent, infrastructur, configur, deploy, cours, vocat |

| | | | |
|---|---|---|---|
| 7 | statistical evaluation of civic education programs | experiment, score, variabl, deviat, quasi, random, femal | Group Africa: neoliber, sweden, korea, volum, freir, paulo, scholarship<br><br>Group Asia: onlin, solidar, australia, internet, crisi, civil, space<br><br>Group co-author: letter, immigr, literaci, spanish, project, write, adult<br><br>Group Eastern_Europe: russia, sociocultur, permit, russian, loyalti, disabl, humanist<br><br>Group Latin_America: devot, upbring, moral, econom, substanti, region, cultiv<br><br>Group Middle_East: declar, arendt, britain, great, distrust, interview, contemporari<br><br>Group west: resist, intervent, literaci, europ, heterogen, twentieth, internationalis |

| | | | |
|---|---|---|---|
| 8 | the quantitative analysis of students' attitudes | regress, efficaci, corrupt, adolesc, latent, confirmatori, explanatori | Group Africa: track, nativ, pillar, contradictori, younger, event, textbook<br><br>Group Asia: polish, holocaust, poland, neutral, childhood, histori, uniti<br><br>Group co-author: vocabulari, realis, czech, section, efficaci, child, pupil<br><br>Group Eastern_Europe: hungari, upbring, participatori, coverag, inequ, pattern, configur<br><br>Group Latin_America: acquisit, higher, specialist, comput, canadian, chines, event<br><br>Group Middle_East: municip, czech, hierarch, presid, select, republ, textual<br><br>Group west: crisi, british, youth, legal, mobilis, conscious, forum |

| | | | |
|---|---|---|---|
| 9 | human rights education | human, right, violat, digniti, convent, amend, freedom | Group Africa: agricultur, scienc, gradual, reform, czech, featur, slovakia<br><br>Group Asia: unemploy, measur, pilot, agricultur, intercultur, entir, cooper<br><br>Group co-author: justif, labor, spatial, readi, canada, internet, nativ<br><br>Group Eastern_Europe: languag<br><br>Group Latin_America: custom, textbook, engin, resist, crisi, disciplin, strengthen<br><br>Group Middle_East: spanish, categori, perform, entir, novel, charter, woman<br><br>Group west: childhood, stori, intercultur, literari, spiritu, ident, indigen |

| | | | |
|---|---|---|---|
| 10 | civic education in Europe | union, european, europ, hungari, cohes, eastern, supran | Group Africa: matrix, immigr, suprem, segreg, index, discurs, climat<br><br>Group Asia: ident, other, environment, ideolog, empathi, engin, unesco<br><br>Group co-author: adult, youth, entir, nineteenth, germani, inspir, represent<br><br>Group Eastern_Europe: colombia, immigr, chapter, aggress, egalitarian, countri, right<br><br>Group Latin_America: resist, impos, career, graduat, object, visual, counteract<br><br>Group Middle_East: paulo, church, latin, cathol, america, brazil, theolog<br><br>Group west: sequenc, detriment, valid, discours, prosoci, chile, discurs |

| | | | |
|---|---|---|---|
| 11 | children and civic education | child, young, childhood, passiv, peopl, heard, famili | Group Africa: program, network, charact, scientist, fiction, access, scienc<br><br>Group Asia: obstacl, mission, prepar, empower<br><br>Group co-author: primari, argentina, languag, onlin, foreign, english, danish<br><br>Group Eastern_Europe: underlin, indispens, greec, legaci, figur, space, regim<br><br>Group Latin_America: reader, chapter, written, literaci, critic, collabor<br><br>Group Middle_East: jewish, nativ, immigr, sector, islam, israel, heritag<br><br>Group west: jordan, candid, holocaust, secur, director, nativ, eighth |

| | | | |
|---|---|---|---|
| 12 | development of civic skills | partner, organ, summer, partnership, leadership, leader, career | Group Africa: justif, scholarship, agent, palestinian, republican, typolog, ideolog<br><br>Group Asia: contextu, episod, eastern, themat, director, jewish, degre<br><br>Group co-author: ireland, elementari, finland, scienc, seventh, stori, english<br><br>Group Eastern_Europe: textbook, mental, academi, detriment, soviet, discurs, historian<br><br>Group Latin_America: ecolog, prospect, curios, justic, northern, injustic, entertain<br><br>Group Middle_East: graduat, operationalis, administ, violenc, habit, deviat, conting<br><br>Group west: alloc, modul, constructivist, viewpoint, turkish, inter, turkey |

| | | | |
|---|---|---|---|
| 13 | global citizenship education | global, decoloni, internation, abroad, unesco, baccalaur, intercultur | Group Africa: lebanon, paint, russian, volunt, transnat, pictur, swedish<br><br>Group Asia: trial, mother, dialog, onlin, event, controversi, refuge<br><br>Group co-author: volunt, foreign, deviat, score, percentag, turkish, variabl<br><br>Group Eastern_Europe: cater, pupil, australian, multilingu, panel, theme, session<br><br>Group Latin_America: digit, detect, agricultur, urban, partner, secur, cooper<br><br>Group Middle_East: civic, palestinian, tension, israel, bilingu, state<br><br>Group west: cathol, worldview, virtu, religion, judgment, religi, intoler |

| | | | |
|---|---|---|---|
| 14 | service learning programs | colleg, servic, presid, public, univers, lectur, academi | Topic 14, Group Africa: african, sociolog, twentieth, historian, black, martin, american<br><br>Topic 14, Group Asia: freir, dissent, agonist, aristotelian, deliber, dewey, crick<br><br>Topic 14, Group co-author: ontario, justic, singapor, preservic, financi, typolog, crisi<br><br>Topic 14, Group Eastern_Europe: drama, theatr, modul, scottish, pupil, operationalis, spatial<br><br>Topic 14, Group Latin_America: peacebuild, refuge, music, japanes, heritag, border, aesthet<br><br>Topic 14, Group Middle_East: holocaust, bulli, simul, latino, prosoci, victim, cohort<br><br>Topic 14, Group west: climat, sport, sociopolit, agricultur, mental, black, counti |

| | | | |
|---|---|---|---|
| 15 | civic literacy | languag, foreign, linguist, learner, speaker, bilingu, sociocultur | Group Africa: court, suprem, technolog, comput, legal, privaci, undocu<br><br>Group Asia: northern, danish, italian, ireland, british, germani, austria<br><br>Group co-author: literaci, medium, preschool, presidenti, onlin, elementari, kindergarten<br><br>Group Eastern_Europe: mathemat, scienc, coach, scientist, inquiri, nonprofit, advocaci<br><br>Group Latin_America: matrix, earth, ecolog, korea, environment, coloni, cosmopolitan<br><br>Group Middle_East: librari, charter, honor, professor, volunt, faculti, graduat<br><br>Group west: canadian, latino, dutch, volunt, immigr, netherland, worker |

4.4 Topic Regional Differentials and Topic Trends

In Figure 10, an upshot of topic trend is provided further by regression analysis. On the horizontal axis is the marginal effect of published year regressed to the annual average topic proportion, whereas the vertical axis represents the average topic popularity in 2020 (the constant term in the regression equation). The horizontal dashed line represents the average of these annual means of topic proportions. We also the statistical significance of the marginal effect under the 95% confidence interval. It can be seen that topics whose marginal effect shows no statistical significance suggest that there are no particular change in topic discussion over the two decades (represented by a circle in Figure 10): these topics are children and civic education, human rights education, and civic literacy. Furthermore, certain topics like measurement of civic education practice, global citizenship education, and statistical evaluation of civic education programs show significant popularity in 2020, yet have negative marginal effect (represented by triangles in the upper left quadrant). This suggests that the interests are high in 2020 but in decline. As shown in the lower left corner, national consciousness and education, the quantitative analysis of students' attitudes, and political economy of civic education are out as shown by negative marginal effect and topic popularity below average. Lastly, the topics with positive marginal effect and below average level of popularity are the up and coming topics: their research interests are on the rise. These topics are civic education in Europe, development of civic skills, curriculum analysis, moral/value education, and service learning programs with service learning program topic being the highest marginal effect. Philosophy discussion shows high marginal effect and high relevance in 2020. Figure 11 also shows the long run trend of topics over time.



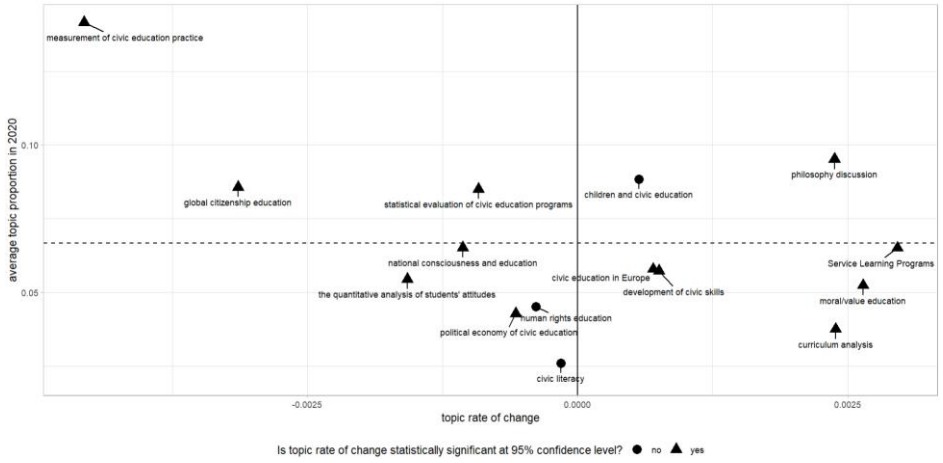


*Figure 10 Scatter plot showing average topic popularity in 2020 and the rate of change in topic popularity*

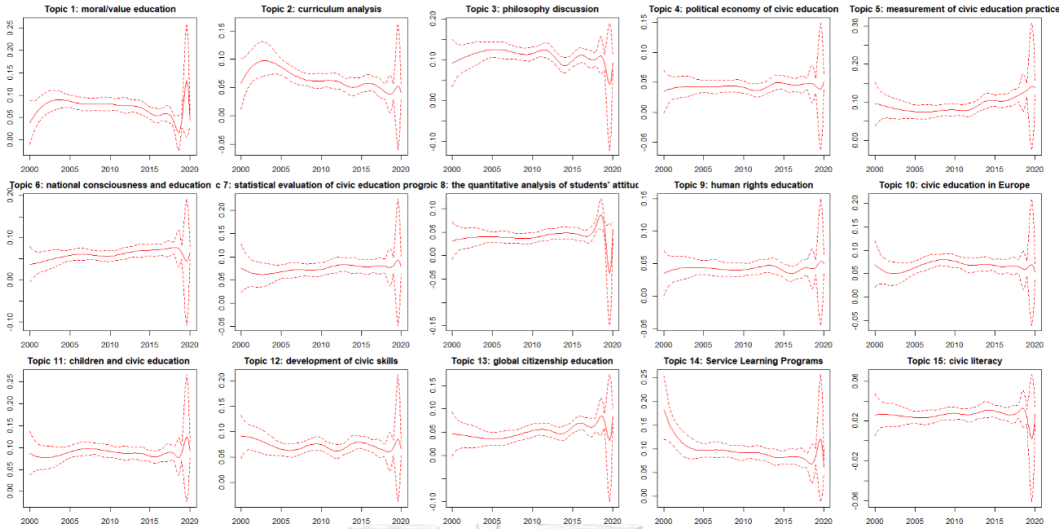


*Figure 11 the long run trend of each topic as generated from the model*

We further analyze the interaction effect between the published year and the regional affiliation in influencing to topic discussion. This is done by decision tree models. The resulted model is as follows. The accuracy of model prediction on the training set and the test set are 13.99% and 14.09% respectively.

From the decision tree result show in Figure 12, we can interpret the interaction between the published year and the regional affiliation as follows. When the published year is between 2000 and 2003, the topic discussion is almost exclusively about topic 5. However, when the published year is between 2004 and 2020, the range of topic discussion varies with the regional backgrounds of researchers. As shown, between 2004 and 2020, Middle Eastern researchers focus on topic 7 while Eastern European researchers concern themselves with Topic 2. Moreover, new change in research interest amongst Western researchers is illustrated: Western researchers whom publish these articles after 2015 are interested more in topic 14 whereas before it was topic 3.

Note that the decision tree model is unable to predict the likelihood of topic discussion for all fifteen topics. This only provides us with a particular interaction between the published year and the regional background of researchers for certain topics. More data is needed to further understand the dynamics of this particular interaction.

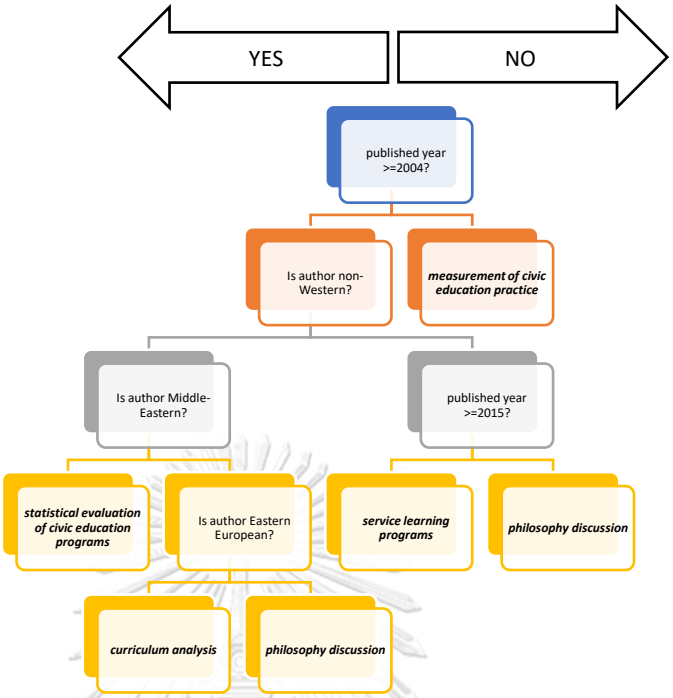


*Figure 12 the chart of decision tree to study the interaction between the published year and the regional background Moving to the left branch is for answer "yes" while going to the right branch is "no"*

Figure 13 shows a tree map of how in each region topics are spread on average across two decades. Figure 14 and Figure 15 show the differential interests amongst researchers based on their regional affiliations. The horizontal axis signifies the topic prevalence contrast or how in a given region each research topic is preferred more over the others. For example, in Figure 14 leftmost subplot, when compared western researchers to their Asian peers they much prefer topic 4 as shown by confidence bands situated in the right part of the subplot. The confidence bands can be interpreted in much like a confidence interval, therefore for the topic bands that go through zero (noted by the dashed line in the middle of each regional graph), it is statistically inconclusive either way. The summaries of this graphical comparison are listed below in Table 4 and Table 5.

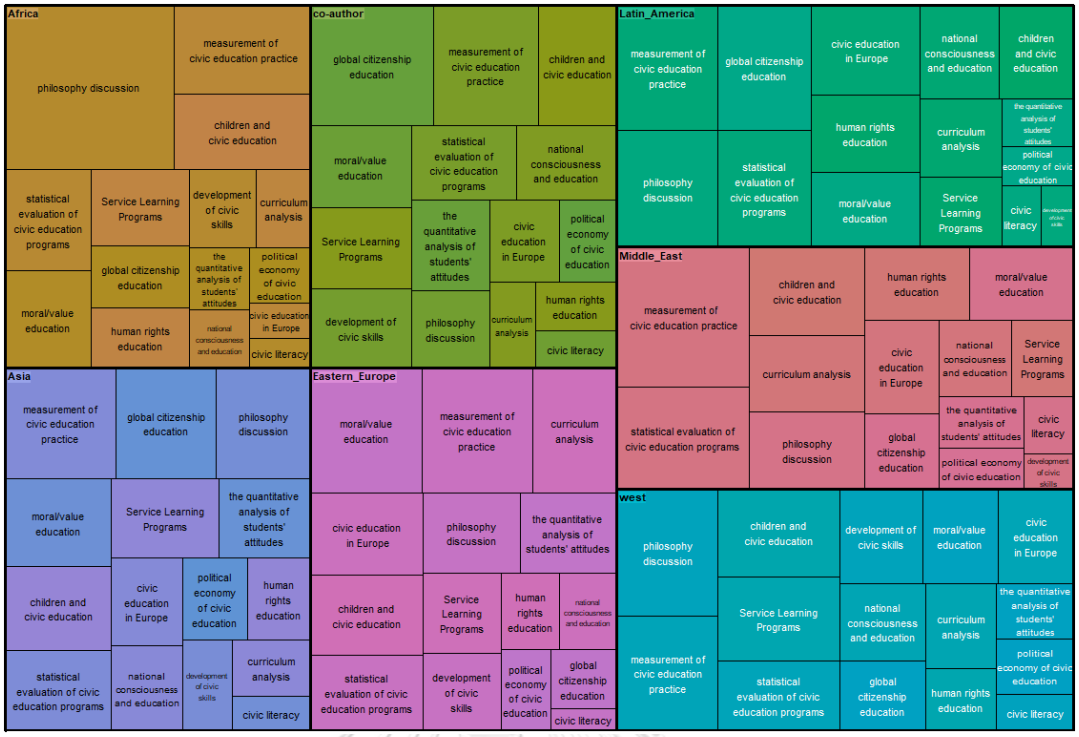


*Figure 13 the treemap showing average topic prevalence in each region. The larger the box, the more prevalent each topic is in the region*

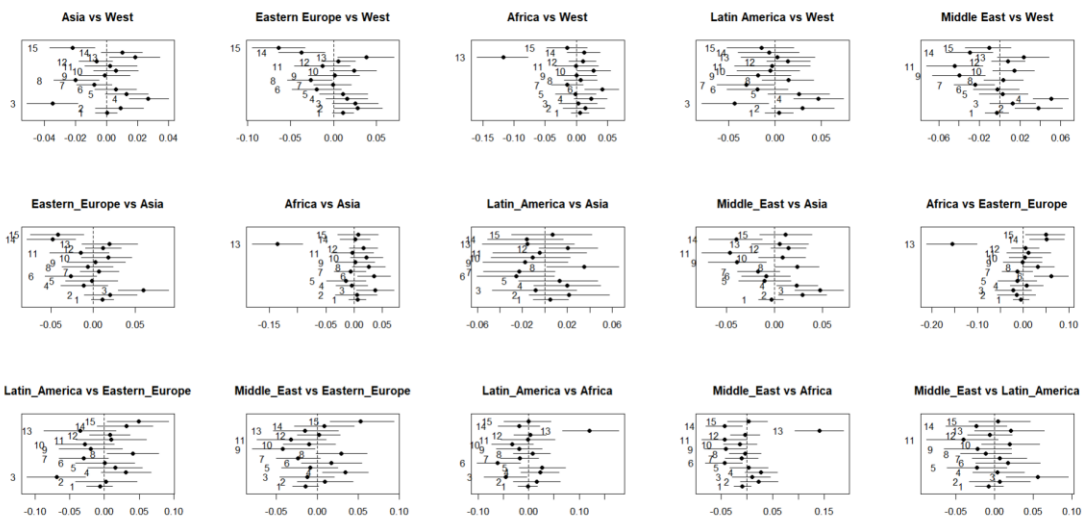


*Figure 14 the topic comparison across different regional background*

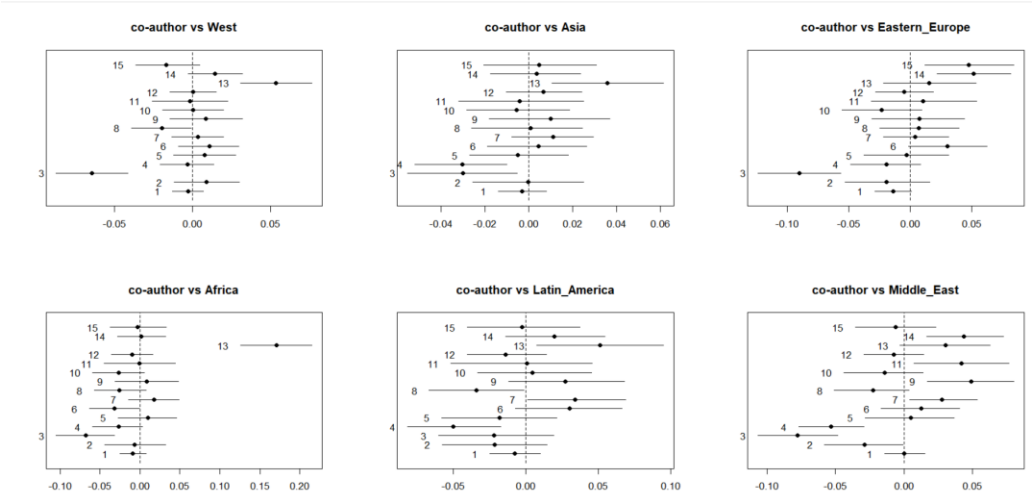


*Figure 15 the topic comparison to the co-author group*

*Table 4: Summary of topics interest based on authors' background*

| *Regional Affiliations* | *Topics of Interest* |
|---|---|
| *Northern America and Western Europe* | global citizenship education (Topic 13)<br>political economy of civic education (Topic 4)<br>curriculum analysis (Topic 2) |
| *Eastern Europe* | service learning programs (Topic 14)<br>civic literacy (Topic 15)<br>the quantitative analysis of students' attitudes (Topic 8)<br>national consciousness and education (Topic 6)<br>political economy of civic education (Topic 4) |
| *Latin America* | philosophical discussion ( Topic 3)<br>statistical evaluation of civic education programs ( Topic 7)<br>national consciousness and education ( Topic 6)<br>global citizenship education ( Topic 13) |
| *Asia* | philosophical discussion ( Topic 3)<br>the quantitative analysis of students' attitudes (Topic 8)<br>global citizenship education (Topic 13) |

| | |
|---|---|
| | civic literacy (Topic 15) |
| *Africa* | global citizenship education (Topic 13) |
| *Middle East* | national consciousness and education (Topic 6) |
| | statistical evaluation of civic education programs ( Topic 7) |
| | human rights education ( Topic 9) |
| | children and civic education ( Topic 11) |

*Table 5: Summary of topics interest based on authors' background as compared to co-authors papers*

| *Regional Affiliations* | *Topics of Interest vs "co-authors"* |
|---|---|
| *Northern America and Western Europe* | global citizenship education (Topic 13) |
| *Eastern Europe* | service learning programs (Topic 14) |
| | civic literacy (Topic 15) |
| *Latin America* | statistical evaluation of civic education programs ( Topic 7) |
| | global citizenship education ( Topic 13) |
| *Asia* | global citizenship education (Topic 13) |
| *Africa* | global citizenship education (Topic 13) |
| *Middle East* | statistical evaluation of civic education programs ( Topic 7) |
| | human rights education ( Topic 9) |
| | children and civic education ( Topic 11) |
| | service learning programs (Topic 14) |

# Chapter 5 Conclusion and Discussion

In this final chapter, we summarize the basic information of our result and provide some discussion as to what this could imply for civic and citizenship education.

## 5.1 Summary of Research

Our topic model analysis suggest there are 15 topics of discussion in our article sample. The highest prevalence of these is research concerning the measurement of civic education, followed by philosophy discussion and the study between children and civic education. These topics are shown to be discussed quite independent of each other as suggested by the low topic correlation.

In terms of trend and changes, our analysis shows that some topics are becoming more popular over time while some are not so. Considering 2020 as our base year, measurement of civic education practice, statistical evaluation of civic education programs, and global citizenship are currently popular but on their way out of popularity. On the other hand, political economy of civic education, national consciousness and education, and the quantitative analysis of students' attitudes are archaic discussion. Philosophy discussion turns to be increasingly famous amongst researchers. Finally, there are a number of up-and-coming topics which are likely to become more focused on by researchers in the future. These are moral/value education, curriculum analysis, civic education in Europe, development of civic skills, and Service Learning Programs.



There are limited interaction between the covariates – the published year and the regional background of researchers – in determining the topics of discussion. As illustrated by our decision tree application, some non-Western researchers 2005 onward tend to prefer certain topics: Middle Eastern researchers prefer statistical evaluation of civic education programs, and Eastern European researchers are more focusing on the curriculum analysis. Western researchers show a recent change in their interest quite recently: after 2015, they tend to prefer research on service learning program.

Global citizenship education is noticeably one of the most popular topics amongst researchers in many countries. African researchers in particular exclusively focus on it as compared

to others. The preceding charts in Figure 16 show discrepancy in topic-covariate discrepancy of the topic where Africa is compared to Northern America and Western Europe, Latin America, and Asia (regions whose researchers are significantly interested in global citizenship education). The comparison suggest although they are interested in the same topic the focus could be slightly different as shown between Latin American and African researchers "biases" towards different words depicted in the graph below.

National consciousness and education (Topic 6) is also another popular topic especially in Middle East, Latin America, and Eastern Europe. When compared the word biases amongst these regions as shown in figure 17 below, the result suggests that Eastern European specifically are interested in Polish history and the Holocaust as suggested by the prominent words. This suggest a strong historical flavor in civic educations studies of this topic in certain regions.

There are two topics which show a clear quantitative flavor: topic 7 and topic 8. Western and African researchers are the only two regions whose topic of interest does not include any of the two topics. Since from our correlation analysis, there does not seem to be any two topics which are usually discussed together. This implied that the interest of global citizenship education by the African researchers may not have been quantitatively oriented. This also may suggest a divide in methodology between researchers in these two regions and the rest.

This is further reinforced when we look at the interests of Western researchers which consist of qualitative discussion like political economy, discourse/content analysis, and global citizenship education. However, when we compare then to the co-author papers, only two regions (Latin America and the Middle East) remain strongly interested in the quantitative analysis.



5.2 Discussion

The trend of research in civic and citizenship education although on the surface seems stable shows ebbs and flow. Topics that show clear flavor in terms of quantitative analysis or politically oriented focus are popular but are shown to be increasingly out of fashion. In turn, topics that concentrate on skill development and educational program construction are on the rise. Nevertheless, the cross-topic research have more room for improvement in the future as our analysis does not seem to suggest codependence of these topics. This is reinforced further by the multi-disciplinary nature of civic and citizenship education

In terms of regional background and its effect, global citizenship education, national consciousness education, and quantitative analysis turn out to be the most popular topics across many regions. However, as we have shown, these topics are on their way out. Service learning program, the only research topic that is shown to be on the rise, is exclusively interested by researchers from Eastern Europe. This is reinforced further when we compare Eastern European researchers to the co-author group as shown. Plus, from our decision tree analysis, Western researchers have also shown to be increasingly more interested in research in service learning programs as well. Philosophical discussion, the common topic in our analysis, seems to be markedly focused on for Latin American and Asian researchers, but these differences disappear when compared to the co-author group.

5.3 Further Suggestion

The result of this research should be used as a way to see in which directions civic education research is heading towards and what is left behind. Although topics like service learning programs and skill development are on the rise, this upcoming trend tends to be exclusively concentrated in certain regions. This encourages researchers from others to delve more into these particular topics in order to engage more in the research of civic education. Moreover, many topics that are “out of fashion” might be worth considering whether why there is oversaturation in research. This might suggest further development in and, in terms of educational policy, allow policy makers to understand further what needs to be focused on to response to changing environment.



More relevant data is required should further research is to examine the discussion in civic and citizenship education. This includes a larger amount of papers to analyze from other different eligible databases and also other background information from both the authors and the published journals. Although there is a vast amount of available data for this particular kind of analysis, restriction on access did hamper on the ability of researchers to gather relevant data. Moreover, we can also study the co-movement of research based on different regional researchers to further understand the global-scale contribution in civic education research.

# REFERENCES


Appiah, K. A. (2006). Global citizenship. *Fordham Law Review*, 75(5). https://ir.lawnet.fordham.edu/flr/vol75/iss5/3/

Apple, M., & Apple, M.W. (2018). Ideology and Curriculum (4th Ed.). Routledge/Taylor & Francis Group.

Bai, X., Zhang, X., Li, K., Zhou, Y., & Yuen, K.F. (2021). Research Topics and Trends in the Maritime Transport: a Structural Topic Model. *Transport Policy*. 10.1016/j.tranpol.2020.12.013.

Bakhtiari, S. (2011). Globalization and Education: Challenges and Opportunities. *International Business & Economics Research Journal (IBER).* 5. 10.19030/iber.v5i2.3461.

Banks, J. (2008). Diversity, Group Identity, and Citizenship Education in a Global Age. *Educational Researcher*. 37. 10.3102/0013189X08317501.

Bannerji, H. (2020). Building from Marx: Reflections on 'race', gender and class. In *The Ideological Condition: Selected Essays on History, Race and Gender* (pp. 5-22). Brill.

Barber, C., Fennelly, K., & Torney-Purta, J. (2013). Nationalism and support for immigrants' rights among adolescents in 25 countries. *Applied Developmental Science*, 17(2), 60-75.

Barde, B.V., & Bainwad, A.M. (2017). An overview of topic modeling methods and tools. 2017 International Conference on Intelligent Computing and Control Systems (ICICCS), 745-750.

Bennion, E. A., & Laughlin, X. E. (2018). Best practices in civic education: Lessons from the Journal of Political Science Education. *Journal of Political Science Education*, 14(3), 287-330.

Bischof, J.M. & Airoldi, E. (2012). Summarizing topical content with word frequency and Exclusivity. *Proceedings of the 29th International Conference on Machine Learning, ICML 2012*, 1. 201-208.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *the Journal of machine Learning research*, 3, 993-1022.

Blei, D. M., & Lafferty, J. D. (2006). Dynamic topic models. *In Proceedings of the 23rd international conference on Machine learning* (pp. 113-120).

Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM*, 55(4), 77-84.

Boontinand, V., & Petcharamesree, S. (2018). Civic/citizenship learning and the challenges for democracy in Thailand. *Education, Citizenship and Social Justice*, 13(1), 36-50.

Boyd-Graber, J. L., Hu, Y., & Mimno, D. (2017). Applications of topic models (vol 11). Now Publishers Incorporated.

Bray, M., & Thomas, R. (1995). Levels of Comparison in Educational Studies: Different Insights from Different Literatures and the Value of Multilevel Analyses. *Harvard Educational Review*, 65, 472-491.

Bureekun, T.,Sanghamad, R. ,& Merieau,E. (2011). ความเป็นพลเมืองในประเทศไทย (Citizenship in Thailand). *King Prajadhipok's Institute*. https://www.kpi.ac.th/knowledge/book/data/740

Castro, A. J., & Knowles, R. T. (2017). Democratic Citizenship Education. *The Wiley handbook of social studies research*, 287.

Carretero, M., Haste, H., & Bermudez, A. (2016). Civic education. In L. Corno & E. M. Anderman (Eds.), *Handbook of educational psychology* (pp. 295–308). Routledge/Taylor & Francis Group.

Chen, X., Xie, H.(2020). A Structural Topic Modeling-Based Bibliometric Study of Sentiment Analysis Literature.*Cognitive computation*, 12, 1097–1129. https://doi.org/10.1007/s12559-020-09745-1

Chi, R., Wu, B., & Wang, L. (2018). Expert Identification Based on Dynamic LDA Topic Model. *In 2018 IEEE Third International Conference on Data Science in Cyberspace (DSC)*, Guangzhou, 2018, pp. 881-888, doi: 10.1109/DSC.2018.00141

Chiva-Bartoll O., Ruiz-Montero P. J., Martín Moya R., Pérez López I., Giles Girela J.,García-Suárez J. y Rivera-García E. (2019). University Service-Learning in Physical Education and Sport Sciences: A systematic review. *Revista Complutense de Educación*, 30(4), 1147-1164. https://doi.org/10.5209/rced.60191

Choi, M. (2016*). A Concept Analysis of Digital Citizenship for Democratic Citizenship Education in the Internet Age*. *Theory & Research in Social Education*, 44:4, 565-607. https://doi.org/10.1080/00933104.2016.1210549

Chow, J. K. F., & Kennedy, K. J. (2015). Asian students' conceptions of future civic engagement: Comparing clusters using person- centered analysis. *Research in Comparative and International Education*, 10(1), 7–22. https://doi.org/10.1177/1745499914567817

Debnarin, P. (2019). ระบบคิดเชิงเหตผลของการศึกษาเพื่อสร้างความเป็นพลเมืองในสังคมไทย.*วารสารการเมือง การบริหารและ กฎหมาย,* 122(1).

Mansup, S. (2020). การศึกษาเพื่อสร้างพลเมือง :ฐานรากของการเมืองภาคพลเมือง. *King Prajadhipok's Institute Journal*,6(2).https://so06.tci-thaijo.org/index.php/kpi_journal/article/view/244664

Morris, P., & Cogan, J. (2001). A comparative overview: Civic education across six societies. *International Journal of Educational Research*. 35. 109-123. 10.1016/S0883-0355(01)00009-X.

Cohen, J. L. (1999). Changing Paradigms of Citizenship and the Exclusiveness of the Demos .*International Sociology*, 14(3), 245– 268. https://doi.org/10.1177/0268580999014003002

Daenekindt, S. & Huisman, J. (2020). Mapping the scattered field of research on higher education. A correlated topic model of 17,000 articles, 1991–2018. *Higher Education*, 80, 571–587. https://doi.org/10.1007/s10734-020-00500-x

Davies, L. (2008) Comparing definitions of democracy in education. In: Arthur, J, Davies, I (eds), *Citizenship Education volume 2: The Purpose of Citizenship Education*. London: SAGE , 138–153.

Davies, M., & Calma, A. (2020). Critical thinking in business education: Current outlook and future prospects. *Studies in Higher Education*.https://doi.org/10.1080/03075079.2020.1716324

Demaine, J. (2002). Globalisation and citizenship education. *International Studies in Sociology of Education*. 12. 10.1080/09620210200200086.

Dieng, A. B., Ruiz, F. J., & Blei, D. M. (2019). The dynamic embedded topic model. *arXiv* preprint arXiv,1907,05545.

DiMaggio, P., Nag, M., & Blei, D. (2013). Exploiting affinities between topic modeling and the sociological perspective on culture: Application to newspaper coverage of US government arts funding. *Poetics*, 41(6), 570-606.

Fillmore, L.W. (1991). When Learning a Second Language Means Losing the First. *Early Childhood Research Quarterly*, 6, 323-346.

Finkel, S., Lim, J., Neundorf, A., Öztürk, A., & Shephard, D. (2022). *Adult civic education and support for democratic values: A systematic review and meta-analysis*. https://doi.org/10.31219/osf.io/427gh

Fitzgerald, J., Cohen, A., Castro, E. & Pope, A. (2021). A Systematic Review of the Last Decade of Civic Education Research in the United States. *Peabody Journal of Education*. 96. 1-12. https://doi.org/10.1080/0161956X.2021.1942703.

Frazer , E. (2008). Key perspectives, traditions and disciplines: Overview. In: Arthur J., Davies I. and Hahn C.(eds), *The SAGE Handbook of Education for Citizenship and Democracy*. London: SAGE, 281–291.

Geboers, E., Geijsel, F., Admiraal, W., & Ten Dam, G. (2013). Review of the effects of citizenship education. Educational Research Review, 9, 158-173.

van Gunsteren, H. (1998). A Theory of Citizenship: Organizing Plurality in Contemporary Democracies (1st ed.). Routledge. https://doi.org/10.4324/9780429502583

Gupta, V., & Lehal, G. S. (2009). A survey of text mining techniques and applications. *Journal of emerging technologies in web intelligence*, 1(1), 60-76.

Grimmer, J., & Stewart, B. M. (2013). Text as data: the promise and pitfalls of automatic content analysis methods for political texts. *Political Analysis*, 21(3), 267–297.

Hahn, C. L., & Alviar-Martin, T. (2008). International political socialization research. *Handbook of research in social studies education*,81-108.

Hickman, L., Thapa, S., Tay, L., Cao, M., & Srinivasan, P. (2022). Text preprocessing for text mining in organizational research: Review and recommendations. *Organizational Research Methods*, 25(1), 114–146.https://doi.org/10.1177/1094428120971683

Hingels, A., Hoskins, B., Manca, R., Mascherini, M. ,& Saltelli, A. (2009). Growing Cohesive Society: The Characterization of Active Citizenship.OECD. https://www.oecd.org/site/progresskorea/44120679.pdf

Kalyanamitra, P., & Bunchanthuek, A. (2019). Civic Education through the Process of Creating a Citizenship in Democracy of Thailand. *SSRN Electronic Journal*. 10.2139/ssrn.3546272.

Kao, A., & Poteet, S. R. (Eds.). (2007). Natural language processing and text mining. *Springer Science & Business Media*.

Kerr, D. (1999). Citizenship education in the curriculum: An international review. *School Field*, 10(3/4), 5-32.

Kerr, David. (2005). Citizenship Education: An International Comparison. https://edudoc.educa.ch/static/xd/2003/59.pdf

Kennedy, K. J. (2012). Asian Students' Citizenship Values and their Relationship to Civic Understanding: An Exploratory Study Comparing Thai and Hong Kong Students. *Research in Comparative and InternationalEducation*, 7(2), 248–259. https://doi.org/10.2304/rcie.2012.7.2.248

Klaiber, J. (2009) The Catholic Church, moral education and citizenship in Latin America. *Journal of Moral Education*, 38:4, 407-420. DOI: 10.1080/03057240903321899

Kuhn, M. (2008). Building Predictive Models in R Using the caret Package. *Journal of Statistical Software*, 28(5), 1–26. doi:10.18637/jss.v028.i05, https://www.jstatsoft.org/index.php/jss/article/view/v028i05.

Mazari, S. M. (2010). Multiculturalism and Islam in Europe. *Policy Perspectives*, 7(1), 91–101. http://www.jstor.org/stable/42909256

Knowles, R. T. (2015). Asian values and democratic citizenship: Exploring attitudes among South Korean eighth graders using data from the ICCS Asian Regional Module. *Asia Pacific Journal of Education*, 35(2), 191- 212.

Knowles, R. T., & Di Stefano, M. (2016). International Citizenship Education Research: An annotated bibliography of research using the IEA ICCS and IEA CIVED datasets. *Journal of International Social Studies*, 5(2), 86-118.

Knowles, R. T., Torney-Purta, J., & Barber, C. (2018). Enhancing citizenship learning with international comparative research: Analyses of IEA civic education datasets. *Citizenship Teaching & Learning*, 13(1), 7.

Kobayashi, V. B., Mol, S. T., Berkers, H. A., Kismihok, G., & Den Hartog, D. N. (2018). Text classification for organizational researchers: A tutorial. *Organizational research methods*, 21(3), 766-799.

Tyson, A. C., & Park, S. (2008). Civic education, social justice and critical race theory. In J. Arthur I. Davies, & C. Hahn(Eds.), *The SAGE handbook of education for citizenship and democracy* (pp. 29-39). SAGEPublications Ltd, https://www.doi.org/10.4135/9781849200486.n4

Lee, J., & Hong, Y. (2014). Business Model Mining: Analyzing a Firm's Business Model with TextMining of Annual Report. *Industrial Engineering and Management Systems*. 13. 432-441. 10.7232/iems.2014.13.4.432.

Lee, H., Kang, P. (2018). Identifying core topics in technology and innovation management studies: a topic model approach. *The Journal of Technology Transfer*, 43, 1291–1317 (2018). https://doi.org/10.1007/s10961- 017-9561-4

Lindstedt, N. (2019). Structural Topic Modeling For Social Scientists: A Brief Case Study with Social Movement Studies Literature, 2005–2017. *Social Currents*. 6. 232949651984650. 10.1177/2329496519846505.

Liu, L., Tang, L., Dong, W. (2016). An overview of topic modeling and its current applications in bioinformatics. *SpringerPlus,* 5, 1608 (2016). https://doi.org/10.1186/s40064-016-3252-8

Manning, N. & Edwards, K. (2014). *Does civic education for young people increase political participation? A systematic review*. Educational Review, 66. https://doi.org/10.1080/00131911.2013.763767

Nadkarni, P. M., Ohno-Machado, L., & Chapman, W. W. (2011). Natural language processing: an introduction. *Journal of the American Medical Informatics Association*: *JAMIA*, 18(5), 544–551.https://doi.org/10.1136/amiajnl-2011-000464

Parker, W. C. (2008). Knowing and doing in democratic citizenship education. *Handbook of research in social studies education*, 65- 80.

Mi, S., Lu, S., & Bi, H. (2020). Trends and Foundations in Research on Students' Conceptual Understanding in Science Education: A Method Based on the Structural Topic Model. *Journal of Baltic Science Education*. 19. 551-568. 10.33225/jbse/20.19.551.

Roberts, M.E., Stewart, B.M., Tingley, D., Lucas, C., Leder-Luis, J., Gadarian, S.K., Albertson, B. & Rand, D.G. (2014). Structural Topic Models for Open-Ended Survey Responses. *American Journal of Political Science*, 58, 1064-1082. https://doi.org/10.1111/ajps.12103

Roberts, M. E., Stewart, B. M., & Tingley, D. (2019). stm: An R Package for Structural Topic Models. *Journal of Statistical Software*, 91(2), 1–40. https://doi.org/10.18637/jss.v091.i02

Rose, M. E., & Kitchin, J.R. (2019). *"pybliometrics: Scriptable bibliometrics using a Python interface to Scopus"*.SoftwareX. 10, 100263

Samudavanija, C. (2014). แนวทางการศึกษาเพื่อสร้างพลเมืองในสังคมไทย. In Samudavanija, C. (Eds). *แนวทางการศึกษา เพื่อสร้างพลเมืองในสังคมประชาธิปไตย (Civic Education)* (pp1-10). Institute of Public Policy Study.

Shi, W.; Lin, Y.; Zhang, Z.; Su, J. (2022) Gender Differences in Sex Education in China: A Structural Topic Modeling Analysis Based on Online Knowledge Community Zhihu. *Children*, *9*, 615. https://doi.org/10.3390/children9050615

Singh, N., Hu, C., & Roehl, W. S. (2007). Text mining a decade of progress in hospitality human resource management research: Identifying emerging thematic development. *International Journal of Hospitality Management*, 26(1), 131-147.

Sumathy, K. L., & Chidambaram, M. (2013). Text mining: concepts, applications, tools and issues-an overview.International Journal of Computer Applications, 80(4).

Westheimer, J., & Kahne, J. (2004). What Kind of Citizen? The Politics of Educating for Democracy. *American Educational Research Journal - AMER EDUC RES J*. 41. 237-269. 10.3102/00028312041002237.

Jiaqi, Z., Kaijun, W., Yunkun,W., Zhongyi, H.,& Hongan, W. (2016) .Mining User-Aware Rare Sequential Topic Patterns in Document Streams. *IEEE TRANSACTIONS ON K B. V*, 28(7). 10.1109/TKDE.2016.2541149

Zhu, J., Snow, K., Kennedy, K., & Mok, M. (2018). Previous civic experience and Asian adolescents' expected participation in legal protest: mediating role of self-efficacy and interest. *Asia Pacific Journal of Education*. 1-18.10.1080/02188791.2018.1493980.

Zheng, B., McLean, D.C. & Lu, X. (2006). Identifying biological concepts from a protein-related corpus with a probabilistic topic model. *BMC Bioinformatics*, 7, 58. https://doi.org/10.1186/1471-2105-7-58

# VITA

| | |
|---|---|
| **NAME** | ปุรณ์ ทองใส |
| **DATE OF BIRTH** | 1 ธันวาคม 2536 |
| **PLACE OF BIRTH** | กรุงเทพมหานคร |
| **INSTITUTIONS ATTENDED** | - BSc Mathematics, Imperial College London<br>- M.Ed Educational Statistics, Chulalongkorn University |
| **HOME ADDRESS** | กรุงเทพมหานคร |