

國立臺灣大學社會科學院經濟學系  
碩士論文

Department of Economics  
College of Social Science  
National Taiwan University  
Master Thesis



課綱與國族認同: 探討 1997 年國中教科書改革的影響

Curriculum and National Identity: Evidence from the  
1997 Textbook Reform in Taiwan

陳煒林

Wei-Lin Chen

指導教授: 林明仁 博士

Advisor: Ming-Jen Lin, Ph.D.

中華民國 106 年 8 月

August, 2017

## 謝辭

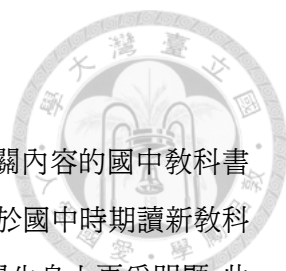
這是一篇關於「認識台灣」這套教科書的文章，研究這個題目的兩年間，我才開始認識台灣。

謝謝林明仁老師鼓勵我將這個題目做成碩士論文，更在我研究陷入低潮時，鼓勵我繼續。謝謝中研院的楊子霆老師，花了無數個下午與我討論這篇文章使用的計量方法和結果詮釋。謝謝研究室 643 的大家，邱錦洋、沈暉智、林建勳、吳東懋、吳愷頡，跟大家一起打屁酸人討論碩論的那段時光，非常開心。謝謝 Prof. Eik Swee, Prof. Martin Huber, Prof. Bei Qin，在研究的初期給予我意見。謝謝江淳芳老師、樊家忠老師、駱明慶老師、Prof. Patrick Dejarnette, Prof. Gerard Roland, Prof. Hans-Joachim Voth, Prof. Noam Yutchman 在這篇文章成形後，給予我非常多寶貴的建議。最後，謝謝我的家人，謝謝爸爸媽媽和姊姊一直相信我，讓我能夠在自己選擇的路上，堅持下去。

這是一篇歪打正著的文章，研究前，我對政治偏好絲毫沒有理解。對於教育體系如何影響人們各式各樣的偏好，亦絲毫沒有想像。然而，現在一旦認真回想義務教育的九年間，那是一個如何封閉且嚴格遵守秩序的社會，不禁讚嘆起，教育真是一部潛在的，嚴謹的洗腦機器。當然，即使真有洗腦這回事，在人們長大後，出教育體系進社會，每天接觸（各式各樣的）資訊，教育體系的洗腦還會有效嗎？被教育體系洗腦的人對社會中其他人的影響是什麼？這篇文章的最後，提出的便是這樣的問題。

關於洗腦這個詞彙，我想補充解釋。通常洗腦帶有負面意涵，但當我使用這個詞的時候，我心裡想的只是一種「偏好的建構」，是中性的。而如果認真回想，你可能會發現，幾乎所有的偏好，某種程度上都是被洗腦出來的，可能是家人、老師、同儕、偉人，也可能是優質電視節目、十大好書、百萬粉絲團。研究洗腦機器的過程中，我愈來愈珍惜言論自由，雖然我們不一定能選擇，但言論自由至少讓我們被洗腦之後，有很多可能的長相。

## 中文摘要



教材能夠改變學生的國族認同嗎？我們檢視了一次引進大量台灣相關內容的國中教科書改革，利用迴歸不連續法去除社會趨勢影響以及世代效果後，我們發現於國中時期讀新教科書的學生更有可能持有較強的台灣人認同。這個現象在就讀學術體系的學生身上更為明顯。此外，藉由分析學生故鄉的族群比例，我們發現如果學生原生故鄉的台灣人認同較低，則教科書的效果較明顯。上述結果隱含兩個教科書影響學生國族認同的管道—背誦教材與鄰里間的社會化。學生年歲較長後，是否曾讀過新教科書變得無法預測學生的國族認同差異。這份結果來自以下事實：未曾讀過新教科書的學生長大後，他們的台灣人認同上升趨勢，比起讀過新教科書的學生更為快速。究竟這份快速的上升趨勢來自於同儕互動中產生的新教科書外溢效果，或是因為整體社會的台灣人認同上升，我們需要更多研究才能回答這個問題。

**關鍵詞：**教科書改革、國族認同、長期效果

## Abstract



Could education content casually affect students' national identity? We exploit the sharp junior high school textbook reform which introduced large amount of Taiwan-related materials, using regression discontinuity design to tease out the society trend and cohort effect. We find that students exposed to the new textbook are more likely to hold stronger Taiwanese identity. The effect is larger for students who entered academic education track and students endowed with less Taiwan-oriented ideology approximated by the ethnic distribution of the hometown, implying mechanisms of memorizing and socialization with neighborhoods. As the students aged, whether they study the new textbook or not could not predict significant national identity differences. We find that this insignificance emerges from the steeper increasing trend of Taiwanese identity in students not exposed to the new textbook. Whether the steeper increasing trend comes from spill-over effect of the new textbook through peer interaction or from the exposure to the also increasing Taiwanese identity trend of the whole society requires further examination.

*Keywords:* Textbook Reform, National Identity, Long Term Effect

# Contents



謝辭	
中文摘要	ii
英文摘要	iii
<b>1 Introduction</b>	<b>1</b>
<b>2 Background and Textbook Analysis</b>	<b>6</b>
2.1 The curriculum reform of <i>Knowing Taiwan</i> series . . . . .	6
2.2 Textbook Analysis . . . . .	7
<b>3 Data</b>	<b>12</b>
<b>4 Main Results</b>	<b>15</b>
4.1 Graphical Analysis . . . . .	15
4.2 Regression Discontinuity Design . . . . .	19
4.3 Assumptions of Regression Discontinuity Design . . . . .	21
4.4 Regression Analysis of the Effect of <i>Knowing Taiwan</i> series . . . . .	24
4.5 Robustness Check . . . . .	29
4.5.1 Bandwidth Choice . . . . .	29
4.5.2 Falsification Test . . . . .	31
<b>5 Subgroup Analysis</b>	<b>33</b>
5.1 Heterogeneity: Education Track . . . . .	33
5.2 Heterogeneity: Hometown Ethnicity Distribution . . . . .	40
<b>6 Long Run Results</b>	<b>47</b>

<b>7 Conclusion</b>	<b>52</b>
<b>A National Identity Distribution and Sample Selection</b>	<b>56</b>
<b>B Observable Characteristics</b>	<b>56</b>
<b>C Bandwidth Sensitivity</b>	<b>60</b>
<b>D Falsification Test</b>	<b>62</b>
<b>E Hoklo Ethnicity Distribution in Taiwan</b>	<b>63</b>



# List of Figures



1	Overall National Identity Trend in the Society . . . . .	17
2	National Identity Trend in Different Education Cohort . . . . .	17
3	Estimated Textbook Effects by Education Cohort . . . . .	18
4	Number of Observations at Month Level . . . . .	22
5	Short Run National Identity Variation at Birth Quarter Level . . . . .	25
6	Bandwidth Sensitivity . . . . .	30
7	National Identity Trend in Different Education Track . . . . .	35
8	Heterogeneity Textbook Effect in Education Track: Cohort Analysis	35
9	Heterogeneity Textbook Effect in Education Track: RD visulization	38
10	National Identity Trend in High/Low Hoklo Ratio Areas in 1990s .	43
11	National Identity Trend in High/Low Hoklo Ratio Areas . . . . .	43
12	Heterogeneity Textbook Effect in High/Low Hoklo Ratio Areas: Cohort Analysis . . . . .	44
13	Heterogeneity Textbook Effect in High/Low Hoklo Ratio Areas: RD visulization . . . . .	45
14	Estimated Textbook Effect by Education Cohort in the Long Run .	49
15	Long Run National Identity Variation at Birth Quarter Level . . . . .	49
16	National Identity Variation by Treatment Status . . . . .	50
18	Observable Discontinuity . . . . .	59
19	Bandwidth Sensitivity with Demographic Controls . . . . .	60
20	National Identity Variation at Birth Month Level with Different Bandwidths . . . . .	61
21	National Identity Variation with <b>Fake</b> Birth Month Cut-Off . . . . .	62





# List of Tables



1	Observables Continuity . . . . .	23
2	Estimated effects of the textbook reform in the Short Run . . . . .	28
3	Falsification Test of Fake Textbook Reform . . . . .	32
4	Heterogeneity Textbook Effect: Education Track . . . . .	39
5	Heterogeneity Textbook Effect: Hometown Ethnicity Distribution . . . . .	46
6	Estimated effects of the textbook reform in the Long Run . . . . .	51
7	National Identity Distribution in Main Regression Sample . . . . .	56
8	Samples Dropped by Sample Selection . . . . .	56
9	Observables Continuity with Polynomials of Second Order . . . . .	57

# 1 Introduction



The more homogeneity the people, the more easiness the management within a nation. Based on this intuition, state leaders have incentives to use education system as an instrument to cultivate national identity, an essential step toward nation building. Empirical evidence has shown that in recent 150 years, the mass education investment by the government appears in response to military threats, when governments acutely requires patriotic people as a preparation for future wars (Aghion et al., 2014). The causal effect underlying the intuition and the transmission mechanism of the effect of education on national identity formation, however, lacks detail scrutinization.

In this paper we investigate the causal effect of education content on national identity formation by examining the junior high school textbook reform in Taiwan. In 1997, the National Institute for Compilation and Translation (*Guoli Bianyi Guan*) published the *Knowing Taiwan* series social subjects textbook. The history volume, in particular, adopted a new perspective of Taiwan's history and provided abundant Taiwan-related content, both of which were absent from previous textbooks. We examine whether studying the *Knowing Taiwan* series causally induce students to hold stronger Taiwanese identity relative to weaker Taiwanese identity.

Using repeated cross section surveys (Taiwan Social Change Survey), we measure the national identity by self reported identity groups and look into the identity variation through both simple cohort comparison and regression discontinuity design (RDD) specifications. We find that students studying *Knowing Taiwan* series are on average 18 percentage points more likely to report themselves as

Taiwanese than Both (Taiwanese and Chinese) and Chinese when students were around their twenties, with a control group mean 0.6.

National identity trend in the society and cohort effects pose challenges to the identification of the education content effect. Specifically, the two effects interact with each other in the sense that students in different cohorts experience different society development. Further, the introduction of *Knowing Taiwan* series may coincide with other social and political events which influence people's national identity. We first compare the national identity in education year cohort level in different survey years to rule out the effect of society atmosphere in different survey years. The national identity variation between the first cohort exposed to the new textbook and the last cohort studying the old textbook is regarded as the textbook effect. To further close out confounding factors that would affect students' national identity *within* an education year, we exploit the identity variation in birth month level. The education system in Taiwan forces children who were born after 1 September to enter next education cohort; namely, children born in August would enter the education system earlier than those born in September for a year. People born in September 1984 would therefore be the first month cohort studying the *Knowing Taiwan* series, while those born in August 1984 would study the old textbooks. The running variable in our regression discontinuity design is thus birth month and the treatment cut-off is September 1984. Thanks to the shifting national identity trend in the society in 1990s in Taiwan, the textbook reform is not exogenous to the society development, but our regression discontinuity design is still valid since the reform is exogenous to the birth month decision of parents giving birth in 1984. Moreover, since people

born near two sides of the cut-off experience very similar society development, the RDD helps us rule out the effect of the society development along our respondents' grown-up.

We examine the possible channel of the *Knowing Taiwan* textbook effect. The effect only appears in students who enter the academic education track when we divide observations by their final education level into vocational and academic tracks. Since students in academic track generally put more efforts on studying academic materials, this result implies that memorizing the education content is a possible channel. Another heterogeneity we investigate into is the prior familiarity of Taiwan-oriented ideology. We approximate the prior Taiwan-oriented ideology with hometown ethnic distribution. Four major ethnicities lived in Taiwan: Hoklo, Mainlanders, Hakka, and Aborigines. In order generations, Hoklo people typically are much more likely to hold Taiwanese identity than other ethnicities. We divide one's living environment by the high/low level of Hoklo people proportion at township level. We find large textbook effect in people who lived in towns with less Hoklo people, whereas no effect is found in people who lived in towns with more Hoklo people. The result is in align with the prediction from "belief-based (bayesian learning) models": people with weaker prior belief are more likely to be persuaded by new information ([DellaVigna and Gentzkow, 2010](#)).

We also study the long run effect of education content and find that people who studied the old textbook hold the Taiwanese identity as strong as those who studied new textbook. This comes from the fact that the increase of the proportion of former group of students who also hold Taiwanese identity is more rapid

comparing to students of the latter group in the long run.

Our paper is relevant to the political preference formation and the culture transmission literature. Previous literature has shown that political regime and media would affect people's policy and party preference ([Alesina and Fuchs-Schundeln, 2007](#); [DellaVigna and Kaplan, 2007](#)) Recently, some researchers examine the relationship between education policy and political institution preference and ethnic identity. [Clots-Figueras and Masella \(2013\)](#) and [Fouka \(2015\)](#) study the language usage promotion and prohibition in education system respectively. [Clots-Figueras and Masella \(2013\)](#) find that the change from single language (Spanish) to bilingual (Spanish and Catalan) education in Catalan makes students have stronger Catalan feelings, and this further leads to changes in political party preference in elections. [Fouka \(2015\)](#), on the other hand, find a backlash effect of German language prohibition in the U.S. education system during 1917 and 1923. She documents that children of German immigrants who experienced the language prohibition are more likely to marry with Germans, choose more German flavor first name for their child, and less likely to volunteer in World War II. She ascribes the backlash effect to the substitute property between parents' investment in children's identity and school's influence on children's identity.

Our paper is most relevant to [Cantoni et al. \(2015\)](#). They conducted a survey in Peking university to study the effect of the new curricula in China which was adopted in different years in different provinces ranging from 2004 to 2010. They find that the new curricula effectively shape students' ideology toward the goal of the curricula; specifically, students exposed to the new curricula exhibit more

trust in government, more skepticism toward unconstrained democracy and unconstrained free markets. They do not, however, find curricular effect on ethnic identity. The goal of the curricula is to make students feel more integrated as a united identity. .



We add up evidence in the direction that other than language usage in education, education content as an education instrument or more broadly a kind of authoritative media, would influence the national identity. Based on the finding of [Cantoni et al. \(2015\)](#), we contribute to the literature by showing that the textbook effect exhibit heterogeneity in the dimension of the effort put into studying the education content and the prior familiarity with the education content. More importantly, the textbook effect might *not* be permanent and we thus have to be cautious about interpretation when we observe the effect of education content policy in the short run. We note that the effect we uncover may include all subsequent behaviors such as the change of media consumption induced by studying new textbooks. We thus interpret the effect as a total impact derived from behaviors entailed by studying *Knowing Taiwan* textbooks.

The paper proceeds as follows: in section 2 we discuss the background of the textbook reform and analyze the difference between the old society subject textbooks and *Knowing Taiwan* series. We describe the data used in this paper in section 3. Main results including graphical analysis and regression analysis based on regression discontinuity design are presented in section 4. Then we explored the potential mechanisms through subgroup analysis in section 5. We explore long run effects in section 6. Finally section 7 concludes.

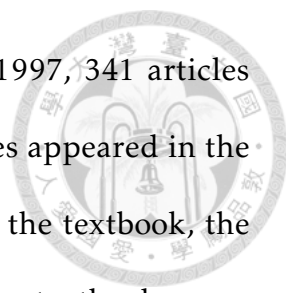
## 2 Background and Textbook Analysis



### 2.1 The curriculum reform of *Knowing Taiwan* series

In 1994, the Ministry of Education of Republic of China announced the new curriculum for junior high school social subjects. The major change lay in the design of the first year content of social subjects. In earlier textbooks, Taiwan-related content accounted for only small portions in the text and was scattered throughout different volumes. The new curriculum of social subjects, including history, geography, and society, aimed to not only provide much more Taiwan-related documentation but also different angles on the history and social development of Taiwan. After three years of writing and editing, the National Institute for Compilation and Translation published the new social subjects textbook, known as the *Knowing Taiwan* series, and students entering junior high school in September 1997 were expected to utilize the series.

The textbook reform aroused fierce debate among political parties on whether the textbook is “appropriate.” Political parties at that time could be divided into two groups. Parties including Kuomintang (KMT) and New Party reckoned the regime as “the successor of China”, while Democratic Progressive Party (DPP) advocated “Taiwan Independence” and considered KMT government, which ruled Taiwan since 1945, as a foreign regime. The discussions at the time about whether the *Knowing Taiwan* history volume should be adopted centered around three perspectives in the textbook: the “relationship between Taiwan and Japan in history,” the “relationship between Taiwan and China in history” and the “judgment of contemporary political events and politicians” (Wang, 2001). According



to Wang (2001), in just two months from June to August in 1997, 341 articles (five articles every day on average) about *Knowing Taiwan* series appeared in the four main newspapers. Facing severe doubt on the content of the textbook, the editors of *Knowing Taiwan* series made minor revisions and the textbooks were still adopted in junior high school across Taiwan in September 1997.

The reform was comprehensive. Students across Taiwan who entered junior high school between September 1997 and September 2000 would study the *Knowing Taiwan* series. Though the major change of the new textbook lay in the ones for the first grade of junior high school. Textbooks for the second and the third year were also adjusted. The senior high school/vocational school entering examinations for students born after Sep 1984 and the earlier education cohort are therefore based on different textbooks of all three years. This makes sure that earlier education cohorts are not exposed to the *Knowing Taiwan* series since they do not have to memorize the materials in the *Knowing Taiwan* series to prepare for the examination.<sup>1</sup>

## 2.2 Textbook Analysis

The curriculum reform aroused politicians' attention because it manifested the stark differences between the two imagined nationalities—the Chinese consciousness and the Taiwanese consciousness (Liu et al., 2005; Wang, 2001). In particular, the new history textbooks divert the angle from “China-oriented” in the earlier textbooks to “Taiwan-oriented”. This is significantly reflected in the nested

---

<sup>1</sup>We define education cohort as students entering compulsory education system in the same year and label them with the year they enter the junior high school. For example, the 1997 education cohort are those who firstly studied *Knowing Taiwan* series. They were born between September 1984 and August 1985. Likewise, the education year refers to the year in education system. The 1997 education year refers to the time period from Sep 1997 to Aug 1998.



variation of the two dimensions between the two versions of the textbooks that being the amount of content of Taiwan history and the context of Taiwan history.<sup>2</sup> The latter could be further divided into two parts: the chronological narrative order and the usage of following terms “our country,” “Taiwan,” and “China.”

Under old curriculum, junior high school students studied the National History for a year and a half and World History for another year and a half, whereas under new curriculum, students studied history about Taiwan in the first year (the *Knowing Taiwan* history volume), National History in the second year, and World history in the third year. In other words, the content of the National History and the World History in the old version is condensed in order to add new materials about Taiwan.

In terms of time, teachers utilizing new textbooks might spend twelve times more on Taiwan history than before. Under old curriculum, only one chapter and a section discuss Taiwan<sup>3</sup>—the third section entitled “The rebellion of Koxinga against Qing Dynasty and the development of Taiwan” in fifteenth chapter in the second volume, and the twenty-fifth chapter entitled “The achievement and vision of a base for revival” in the third volume. Assuming that teachers spend the same time on each chapter and section in a volume, we approximate that teachers would spend less than fifth of a semester on Taiwan history. In contrast, the *Knowing Taiwan* history volume was designed to cover two semesters. The *Knowing Taiwan* history volume has 116 pages about Taiwan. The older textbooks contained only sixteen pages.

---

<sup>2</sup>Taiwan history is referred to as history about Taiwan “island,” instead of the one about Republic of China

<sup>3</sup>In the old textbook series, National History consisted of twenty five chapters in three volumes, each volume for a semester

The notable increment of Taiwan history showed that the editors put more emphasis on it. As the third point of editors' preface in the *Knowing Taiwan* writes:



This book aims to introduce students to *know the history about how ancestors of different ethnic groups made developments in Taiwan*. By learning this, students will cultivate cooperation spirit, patriotic feelings, and worldwide horizons. Also, this will augment students' understanding of *Taiwanese cultural assets*, and make them appreciate and treasure it.<sup>4</sup>

The intention of making students acquainted with Taiwan development is never seen in the old version. We contrast this by recording the editors' preface in the National History in the old version below:

The National History describes *the evolution of Chinese nationality, the change of the territory, and the development of politics, society, economics, and culture*. In particular, it stresses the long history and the blending of the culture of nationality in order to strengthen the patriotic feelings and cooperation spirits, and to know the nation's traditions, the position and the responsibility of national people.

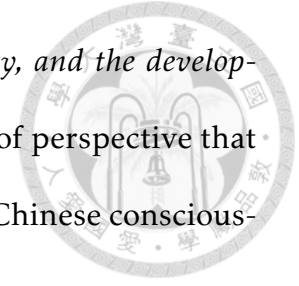
In the new version of the National History textbooks, approximately the same words appear in the editors' preface except that "the evolution of Chinese nationality, the change of the territory" part is deleted.

Obviously, the editors of the *Knowing Taiwan* history volume did not incorporate the *development in Taiwan and the Taiwanese cultural assets* in the context

---

<sup>4</sup>Emphasis of the quote from the textbooks in this section is added by the author

of the evolution of Chinese nationality, the change of the territory, and the development of politics, society, economics, and culture. It is this change of perspective that highlights the battle between “Taiwanese consciousness” and “Chinese consciousness.”



That editors in the new version viewed Taiwan history as a distinct entity different from China history is as mentioned before denoted with the narrative order and the usage of “our country,” “Taiwan,” and “China.”

In their first grade, junior high students who studied the old textbooks started learning the history of “our country” with the statement that the earliest human beings lived in “our country” is *Homo erectus pekinensis* in Paleolithic age. The “common ancestor” of *Chinese nationality* is Huang Di. The first dynasty of “our country” is Xia dynasty. The history of “our country” then developed under sequential dynasties, from Xia to Qin, Tang all the way to Qing. Between the dynasty of Ming and Qing in this straightforward development line rests the first appearance of “Taiwan,” identified by editors as a base for Koxinga to fight against Qing regime.

Despite just a verb, it is worth noted that Koxinga is written as “recovering” Taiwan from the Dutch. The usage of the verb explicitly demonstrates the ideology behind the old textbook, showing that the editor reckons the ruling power of the Dutch in 17<sup>th</sup> century as a “foreign regime.” Simultaneously, this implicitly claimed Taiwan as the territory of “our country” before the ruling of the Dutch.

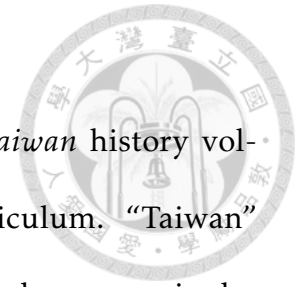
At the end of the line is another appearance of Taiwan, recognized as the revival base under the development of Republic of China. In the new curriculum, however, this straight line of “our country” is decomposed into two separate

strands: Taiwan and China.

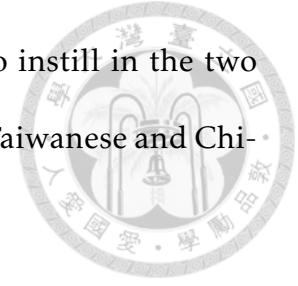
The word “our country” is rarely used in the *Knowing Taiwan* history volume and the National history textbooks under the new curriculum. “Taiwan” and “China” filled the blanks. More precisely, “our country” only appears in description about Republic of China. Following the divided usage of terms, Taiwan history stands out not as part of China history, but as an individual entity in the *Knowing Taiwan* history textbook, and the growing background of Taiwan is introduced: *the history about how ancestors of different ethnic groups made developments in Taiwan*.

Developments in Taiwan are stated in order. Cultures in prehistory era and the aborigines are stated as the first group of people living in Taiwan. What follows are the trade activities among pirates from China and Japan in 16<sup>th</sup> century, and the governing of the Dutch and the Spain. Ruling of Koxinga, Qing dynasty, Japan, and finally the Republic of China are all documented as developments of different groups. On the other hand, the National history textbook under the new curriculum maintains most of the content and the narrative order of the old textbook.

The battle between Taiwanese consciousness and Chinese consciousness would never emerge in the old curriculum since “Taiwan” is almost neglected and “our country” is the only imagined nation. Only when “Taiwan” and “China” are explicitly separated do readers have the chance to distinguish concepts between them. The *Knowing Taiwan* history volume may cultivate Taiwanese identity in two ways. First, priming effect arises from students reading the word “Taiwan” more often. Second, the distinction made by describing Taiwan history and China



history separately may provide students different materials to instill in the two imagined groups and hence helps them differentiate between Taiwanese and Chinese.



### 3 Data

#### The TSCS and the measure of national identity.

The Taiwan Social Change Survey (TSCS) provides repeated cross-section representative observations across Taiwan. It holds two surveys of different topics every year and covers observations aged above 18, each with sample size about 1800 to 2200. The first education cohort exposed to the *Knowing Taiwan* reform, who were born after Sep 1984, is firstly surveyed in 2003. To balance the observations before and after the reform in regression analysis, we hence include surveys held after 2003 which contain the national identity questions and enough background information, including 2003, 2004, 2005, 2010, 2012, 2013 and 2014 waves.

The national identity is measured by the following question:

- In our society, somebody call themselves Taiwanese, somebody call themselves Chinese, and somebody call themselves Both. Do you consider yourself as Taiwanese, Chinese, or both?

We create the dummy variable *Identity* by assigning one to respondents answering Taiwanese and zero to those answering Chinese and Both. We drop the observations who answers "Other." In our main regression sample, only 3.84% of respondents answer "Chinese," indicating that in this generation, very few people

hold pure Chinese identity. We hence adopt the above categorization to differentiate the inclination of holding stronger or lesser Taiwanese identity.<sup>5</sup>

Since the measurement of the national identity is based on the self-reported response, the natural question appears: could this measurement truly reflect respondents' national identity? One possible explanation of the change of *Identity* (if it is observed) is that previous students were afraid of responding himself as Taiwanese. The new textbooks provide not the Taiwanese identity, but the message that saying oneself is Taiwanese is not a taboo anymore. We provide two counterarguments of this explanation. First, the simple mean of *Identity* of the control group in our main analysis sample is 0.6. When over half of one's peer would identify themselves as Taiwanese, it is hard to believe that Taiwanese identity is a taboo for him. Second, the change of *Identity* should be spot in different subgroups under this explanation, but in section 5 we would find that this is not the case.

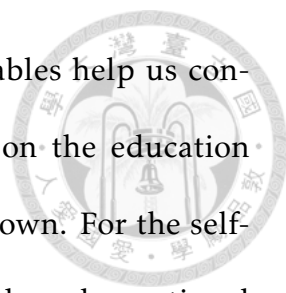
### **The Advantage of TSCS.**

Two features of TSCS make it suitable for our analysis. First, TSCS records the birth year *and* birth month of respondents. This allows us to identify correct education cohort and makes feasible the regression discontinuity design on month level, letting us look into finer variation between birth months instead of yearly cohorts. The second feature of TSCS is the representative sampling and rich demographic variables. The representative sampling allows us to first estimate the treatment effect in the whole population and then further investigate the mechanism through subgroup analysis.

We include the respondent's gender, their parents' education level and their

---

<sup>5</sup>We document detailed proportion of respondents answering different options in Appendix A



parents' ethnicity in the analysis. The rich demographic variables help us construct different subgroups. In particular, we put emphasis on the education track and the ethnic group distribution in respondents' hometown. For the self-education subgroup, we split the sample into academic track and vocational track. We construct *Edu-Academic* dummy variable by assigning one to the respondents whose final education attainment is senior high school or university (academic track); zero to those with junior high school, senior vocational school, college and tech university (vocational track).

### **The proxy of hometown ethnicity distribution.**

One of the questions in TSCS reads: "Where do you live longest before 15 years old?" The response is in township (zip code) level. We regard the response as where the respondents live when in junior high school, and regard it as his hometown.

The hometown information is combined with ethnicity township level data to approximate how many Hoklo people surrounded the respondents in his daily life when in junior high school. The ethnicity data is from National Hakka Population Basic Information Survey Research conducted in 2004 with sample size 37693, about 100 in a town. We use the response of the question: "You consider yourself as?" Six options include: (1) Taiwan Hakka, (2) Mainland Hakka, (3) Hoklo, (4) Mainlanders, (5) Aborigines, (6) Foreigners. The respondents could only pick one answer to this question. The proportion of people answering Hoklo, for example, would be regarded as the proportion of Hoklo people in the town. We then estimate the population median of Hoklo people ratio in Taiwan using Hoklo people proportion in each town and the population of the town in 2004 as

weights (population data from the Ministry of Interior). The median is 77.1%.<sup>6</sup> Finally we construct the dummy variable *Home-Hoklo-Ratio* by assigning one to the respondents whose hometown's hoklo people proportion is less than 77.1%, and assign zero otherwise. If the respondent lived in a town with *Home-Hoklo-Ratio* as one, he is less likely to meet a Hoklo person in his daily life comparing to an average person in Taiwan.

### **Sample Selection Rule.**

We drop the respondents who reports himself born outside Taiwan and the respondents who reports himself live longest outside Taiwan before 15, since we could not make sure that these observations entered junior high school in Taiwan and hence were exposed to the textbook reform. We also drop respondents who did not report their parents' ethnicity and education level, which are required demographic information we need for the regression analysis. Finally, we drop respondents whose answer of the national identity question is "Other." These selection rules drop 19 (4.5%) of the main regression sample (the main regression sample size is 417).<sup>7</sup> The main results in this paper are not influenced by the sample selection.

## **4 Main Results**

### **4.1 Graphical Analysis**

Our objective is to investigate the relationship between studying *Knowing Taiwan* series and holding stronger Taiwanese Identity. The design of the textbook reform

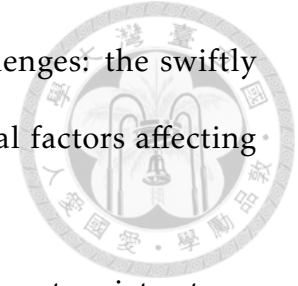
---

<sup>6</sup>The mean using the same data and weight is 73.2%

<sup>7</sup>We provide detail numbers of dropped observations in each step in Appendix A for all samples we use in this paper.



and the repeated cross section survey data bring us two challenges: the swiftly altering national identity trend in the society and the potential factors affecting national identity formation within an education cohort.



Younger cohorts are more likely to be influenced by more recent society atmosphere, which is more Taiwanese-oriented. An obvious increase of the proportion of people responding himself as Taiwanese in the last two decades is spot in Figure 1, which depicts the national identity trend derived from observations across all ages and measured by National Cheng Chi University Election Study Center.<sup>8</sup> To control for this channel and the survey year fixed effect, we stratify the observations in education-cohort-survey-year level. We include samples with education cohorts ranging from 1992 to 1998 and then investigate the national identity variation at education cohort level. The textbook reform changes the treatment status (studying certain textbooks) on an education cohort basis. Figure 2 plots the simple mean of *Identity* in each cell. The sample size in each cell is on average 41.

Two noticeable facts could be seen in Figure 2. First, we observe a jump with roughly 0.2 magnitude between the 1996 and 1997 education cohort, the first cohort studying *Knowing Taiwan* series. To examine if the jump is statistically significant, we regress *Identity* on education cohort and survey year dummies using 1996 education cohort as the reference group. We plot the point estimates and 90% confidence intervals of education cohort dummies in Figure 3. The point estimates of previous education cohorts are not significantly different from 1996 education cohort. The 1997 education cohort dummy is significantly dif-

---

<sup>8</sup>We do not use the data from National Cheng Chi University Election Study Center due to data accessibility and the lack of birth moth records in the data.

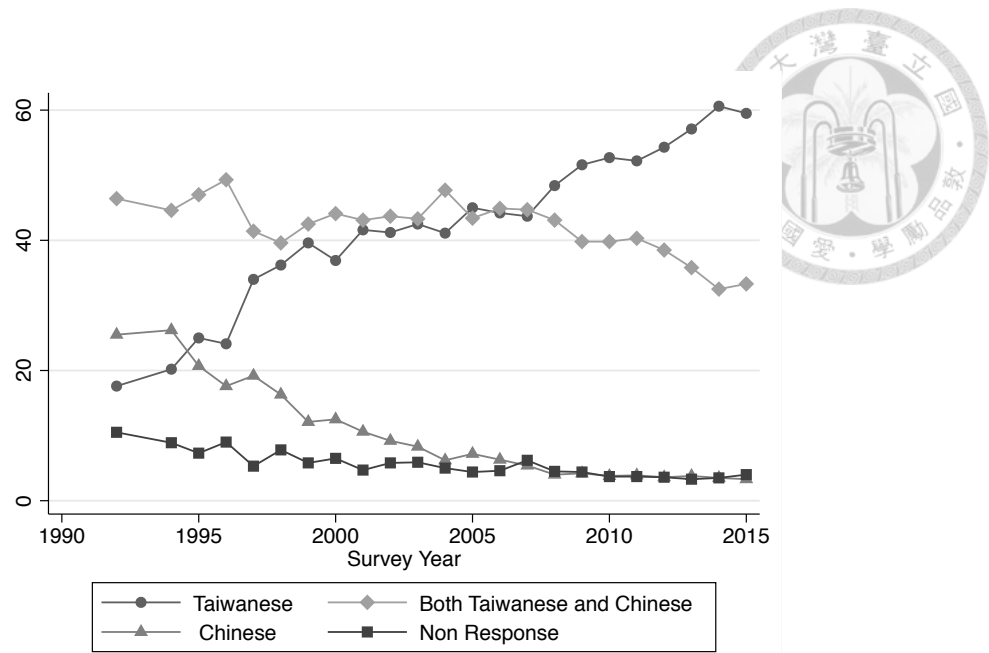


Figure 1: Overall National Identity Trend in the Society

Notes: Cheng Chi University Election Study Center conducts telephone survey on adults aged 20 above every year. The sample size ranges from 1,200 to 34,000; on average it is about 13000. The question used to measure the national identity adopted by Cheng Chi University Election Study Center is the same as the one in TSCS.

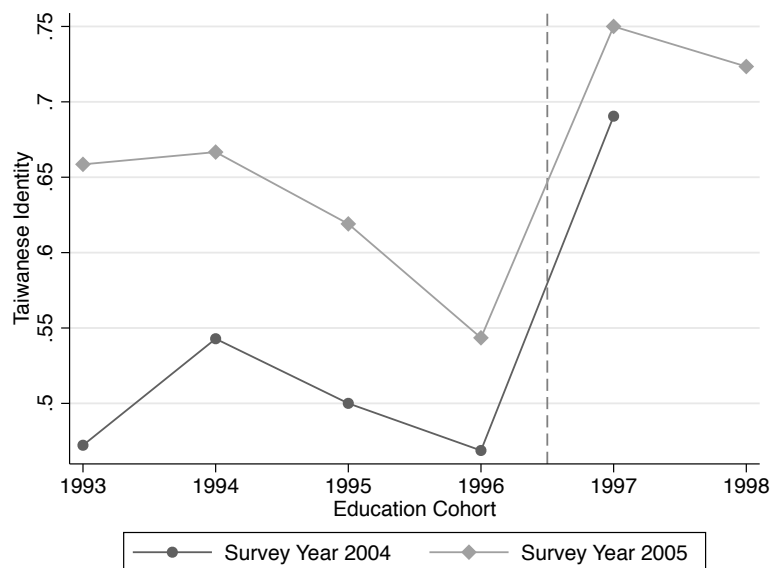


Figure 2: National Identity Trend in Different Education Cohort

Notes: Data is collapsed at education cohort-survey year level from 2004 and 2005 wave of TSCS. Dots represent the simple mean of *Identity* at each cell. We do not include the cell mean derived from 2003 survey year since the sample size for the treatment cohort is too small (less than 20). For the same reason, we drop the mean of 1998 education cohort surveyed in 2004.

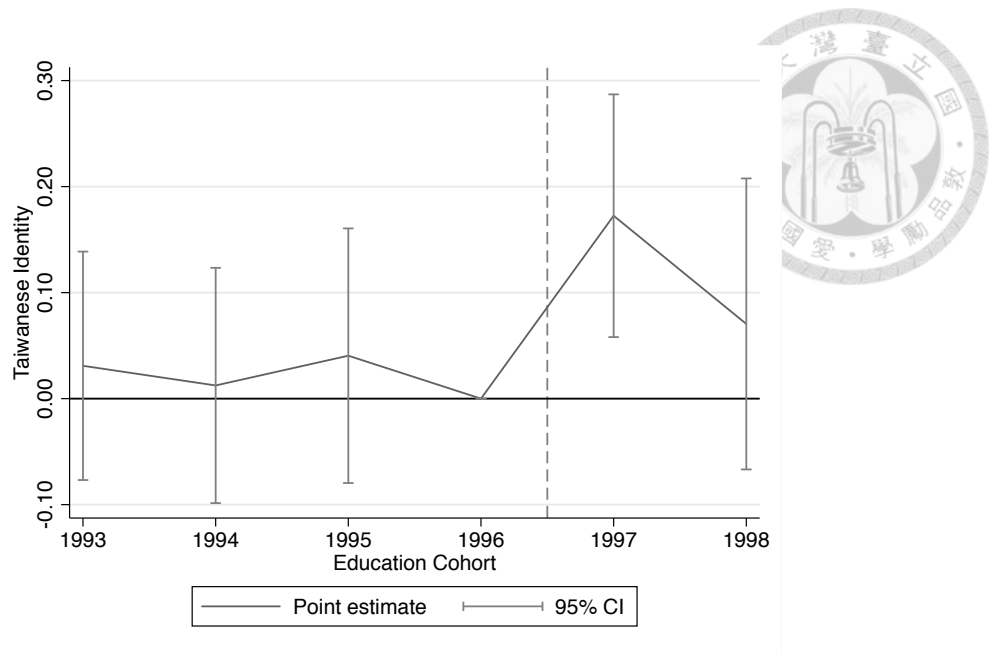


Figure 3: Estimated Textbook Effects by Education Cohort

Notes: We pool 2003, 2004, 2005 TSCS waves. Point estimates derived from regression of *Identity* on survey year dummies, hometown dummies and education cohort dummies (1996 as reference group). Standard errors are clustered at birth month level.

ferent from the 1996 education cohort, with magnitude 17.3 percentage points. The insignificance of 1998 education cohort dummy may arise from the small sample size of the 1998 education cohort. This is the first evidence showing that the treatment, studying *Knowing Taiwan* series, causes students to hold stronger Taiwanese identity.

The second important observation from Fig 2 is that the survey year fixed effect could not capture the upward time trend of national identity fully. This could be seen by observing that the difference between two lines from two survey years are not the same across education cohorts. To avoid this problem, we further restrict our sample to fewer education cohorts in the rest of our analysis: observations who were born between Sep 1982 and Aug 1986 (four education cohorts, two studied *Knowing Taiwan* series). We first analyze these cohorts when they are relatively young, aging from 18 to 23 and surveyed from 2003 to 2005

(henceforth **short run**). To examine if the textbook effect is persistent, we examine the same education cohorts surveyed from 2010 to 2014, with age range 24 to 32 (henceforth **long run**).



## 4.2 Regression Discontinuity Design

The society trend problem prompts us to compare observations who experience almost the same society development along their grown-up. Regression Discontinuity Design (RDD) enables such task. At first glance, we should conduct RDD on education cohort basis since the treatment status varied at that level. Nevertheless, people in the same education cohort may experience different events possibly altering their national identity, for instance, the voting history.

Elections in Taiwan are generally held in December, January, and March. The eligible age is 20. In some elections, people born in the first half of education cohort are eligible, while people in the second half are not. Students who were the first cohort exposed to the textbook reform offers one example. The sixth legislative election was held on Dec 11, 2004, splitting the education cohort into two groups: people who have voting right (born before Dec 11, 1984) and people who do not (born after Dec 11, 1984). Students who were the last cohort studying textbook of the old version provides another example. The election splitting their vote history is the president election in Mar 20, 2004. Since politicians in Taiwan debate fiercely on the national identity issue in elections, the different "first vote" experience may in turn affect people's national identity formation. Bearing in mind such differences embedded with respondents within an education cohort,

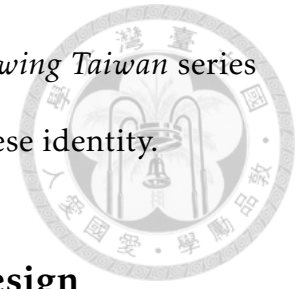
we adopt the standard RDD specification at the birth month level:

$$Identity_{ijt} = \alpha_0 + \alpha_1 TextbookExposure_i + f(m) + \gamma X_i + \eta_j + \delta t + \epsilon_{ijt} \quad (1)$$

where  $Identity_{ijt}$  indicates the dummy variable defined in the data section of individual  $i$ , with hometown  $j$  interviewed at time  $t$ . The variable  $TextbookExposure$  indicates whether the respondent was exposed to the textbook reform and takes 1 if the respondent reports himself born after Sep 1984, 0 otherwise. We use birth month as our running variable and recenter the birth month at Sep 1984, the first month cohort affected by the reform. Following standard practice, we include  $f(m)$ , a smooth function of birth month, to control for the impact of society trend. In our case,  $f(m)$  is first order or second order polynomials fully interacted with  $TextbookExposure$ . Standard errors are clustered at birth month ( $m$ ) level. The month level RDD provides a stronger identification than the cohort comparison in Fig 3 since people around the cut-off should experience very similar society atmosphere. The only disparity around the cut-off is the entering year of compulsory education, which subsequently includes our treatment, the studying of *Knowing Taiwan* series in junior high school.

To increase the precision of our estimates, In all specifications, we include survey year fixed effect ( $\delta_t$ ) to single out overall effect of the society trend in each survey year. we also include hometown fixed effect ( $\eta_j$ ), which help us control for regional factors possibly influencing the national identity formation, e.g., the local support for certain political party. In some specifications, we include demographic variables ( $X_i$ ) which might influence the national identity formation, including gender, parent's education, parents' ethnicity, and *Home-Hoklo-Ratio*.

The core interest is  $\alpha_1$ , which indicates whether studying *Knowing Taiwan* series would causally make one more likely to hold stronger Taiwanese identity.



### 4.3 Assumptions of Regression Discontinuity Design

The validity of the RDD rests in the assumption that no precise control exist at the cut-off. Specifically, if the “quarter of birth” is endogenous in Taiwan, the assumptions of the regression discontinuity design may be violated. [Fan et al. \(2014\)](#) shows that contrast to United States, high socioeconomic status mothers have similar birth pattern with low socioeconomic status mothers, assuring us that birth seasonality is not a concern in our paper. We then check the smoothness of the number of observations and conduct standard practices of regression of observables to validate our empirical strategy.

#### The Smoothness of Observation Frequency.

We use survey data in this paper, so we could not detect if the population around the cut-off exhibit discontinuity. Instead, we plot the number of observations in each month cell to see if the survey design give more weight to one side of the birth month cut-off. As [Figure 4](#) shows, discontinuity is absent. The discontinuity at the fourth month after the cut-off arises due to the fact that TSCS only surveyed adults aged 18 above, so people who were born in 1985 would not be surveyed until 2004.

#### Smoothness of Observables Mean.

To test if the distributions of the observables are smooth around the cut-off, we specifically run regression on observables related to national identity using specifications as [equation 1](#), with first order birth month polynomial. Regression

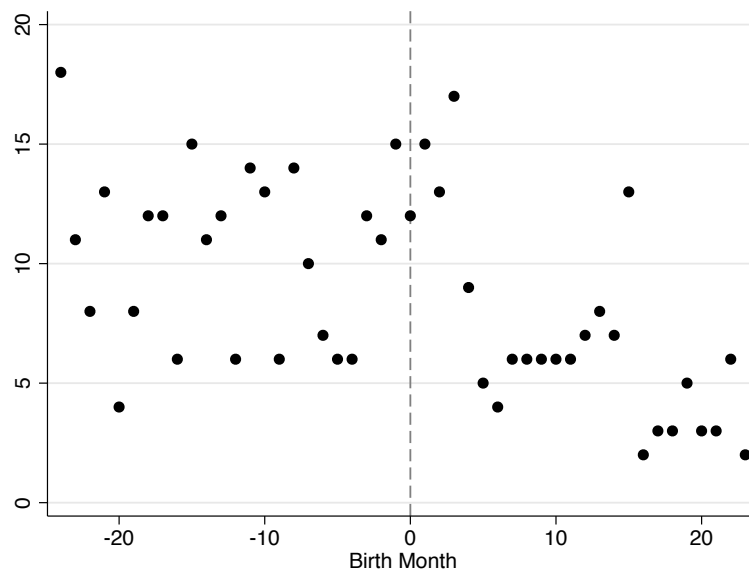


Figure 4: Number of Observations at Month Level

*Notes:* I pool data from 2003, 2004, 2005 TSCS and include education cohorts from 1995 to 1998. Each dot represent the number of observations in that birth month. The running variable (birth month) is recentered at Sep 1984, the first month cohort expose to the textbook reform.

results are shown in Table 1. The counterpart graphs are in Appendix B.

*Home-Hoklo-Ratio* exhibits significant discontinuity with size 0.146 in Table 1, showing that it is significantly more likely to observe a respondent who had lived in a town with less Hoklo people on the right hand side of the cut-off. We argue that this discontinuity comes from sampling error. Since we have few samples in each birth month cell, it is likely that due to sampling error the survey data we used display different characteristics from the population . One evidence supporting our argument is that when we run the same regressions in the long run sample, no significant jump is found. We include the *Home-Hoklo-Ratio* variable into some of the specifications to lessen the potential bias. For other observables, we do not find discontinuities in Table 1.

Table 1: Observables Continuity

VARIABLES	gender	Father Ethnicity	Mother Ethnicity	Father Edu Level	Mother Edu Level	Father Edu Years	Mother Edu Years	Self Edu Level	Self Edu Years	Hometown Hoklo Ratio
<i>TextbookExposure</i>	0.0398 (0.0713)	-9.30e-05 (0.0777)	0.0316 (0.0569)	-0.0747 (0.0897)	-0.0265 (0.0837)	-0.0583 (0.605)	-0.0242 (0.447)	-0.0316 (0.0767)	-0.0641 (0.251)	0.146* (0.0736)
<i>Month</i>	0.000844 (0.00369)	0.00222 (0.00401)	0.000127 (0.00403)	0.00110 (0.00411)	-0.000559 (0.00351)	0.0268 (0.0326)	0.0225 (0.0224)	0.00316 (0.00307)	-0.0275* (0.0152)	-0.00417 (0.00403)
<i>TextbookExposure</i> <i>×Month</i>	-0.00649 (0.00664)	-0.00732 (0.00643)	-0.00472 (0.00517)	0.00641 (0.00770)	0.00247 (0.00692)	-0.0246 (0.0485)	-0.0207 (0.0384)	-0.00724 (0.00609)	0.0376 (0.0247)	-0.00242 (0.00538)
Constant	0.459*** (0.0582)	0.260*** (0.0519)	0.174*** (0.0459)	0.542*** (0.0636)	0.453*** (0.0382)	10.76*** (0.527)	9.955*** (0.238)	0.564*** (0.0538)	13.90*** (0.171)	0.379*** (0.0647)
Observations	417	417	414	417	417	417	417	417	417	417
R-squared	0.002	0.004	0.003	0.004	0.001	0.004	0.003	0.003	0.012	0.007

Notes: We pool data from 2003, 2004, and 2005 TSCS waves and include education cohorts from 1995 to 1998.. We run regression with specification 1. First order polynomial is included and we do not include demographics. *TextbookExposure* is 1 if the birth month of the respondent is after Sep 1984, 0 otherwise, indicating if one studies the *Knowing Taiwan* series. *Month* is the birth month recentered at Sep 1984. Standard errors are clustered at birth month level in parentheses \*\*\* p<0.01, \*\* p<0.05, \* p<0.1

The construction of the independent variables are the following: Gender: female being 1, male being 0. Father/Mother ethnicity: Hoklo fathers/mothers being 0, otherwise 1. Father /Mother education level: father/mother with no education, elementary school, and junior high school education level being 0, otherwise 1. Father/Mother/Self education years is discrete variable ranging from 0 to 16. Self education level: respondents with final education attainment as senior high school and university being 1, otherwise 0. Hometown Hoklo Ratio is the dummy variable *Home-Hoklo-Ratio*, see construction in section 3 **The proxy of hometown ethnicity distribution.**





#### 4.4 Regression Analysis of the Effect of *Knowing Taiwan* series

We focus on the Taiwanese identity variation measured by the dummy variable *Identity* around the birth month cut-off. We show smoothed figures to examine discontinuities around cut-off and present the estimates derived from equation 1. Unlike standard RDD, plotting simple mean of *Identity* in each birth month cell could not tell apart the survey year fixed effect and the treatment effect, leading to misinterpretation. Also, since the sample size of each birth month cell is too small (9 observations on average), all of the following graphs are in birth quarter level to visualize more informative variation. Aiming to control for the survey year fixed effect, in Figure 5, we follow Shigeoka (2014) and plot the cell mean at birth quarter level of the residual of *Identity* after being regressed on survey year dummies. The lines in Figure 5 represent fitted regressions of the cell mean dots using first order polynomials interacted with the dummy variable *TextbookExposure*.

Although we do not observe a distinct jump of the dots just around the cut-off, the fitted line suggests that the discontinuity of *Identity* is roughly 20% around the cut-off. By comparing the near four dots (an education cohort) on both sides of the cut-off, we notice that the proportion of people holding stronger Taiwanese identity significantly increases. It should be noted that the interpretation of the last two dots on the right hand side of the cut-off, which represent the latter half of 1998 education cohort, requires further caution, since the observations in each dot is only 11.

Table 2 shows the regression results of estimation of specification 1. Hometown and survey year fixed effects are included in all regressions and we find

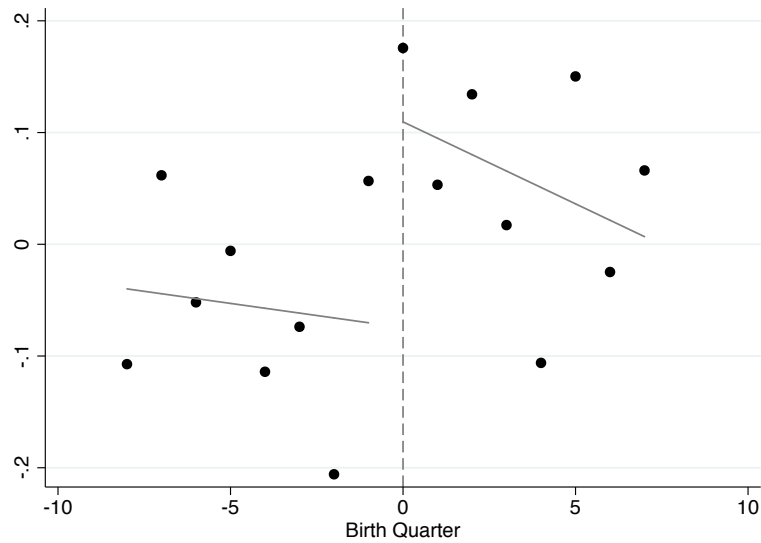


Figure 5: Short Run National Identity Variation at Birth Quarter Level

Notes: I pool data from 2003, 2004, 2005 TSCS and include education cohorts from 1995 to 1998. We first regress *Identity* on survey year dummies and then collapse the residuals at birth quarter level to derive the dots. Fitted lines are from regression of the dots on a first order polynomial of birth quarter interacted with *KnowingTaiwan* dummy variable.

qualitatively similar estimates of the treatment effect across different specifications. Column [1] reports the regression results with first order polynomials controlling the time trend. The estimates of *TextbookExposure* is 0.18 and significant, signifying that studying *Knowing Taiwan* series on average increase one's probability of reporting himself as Taiwanese by 0.18, a much more stronger Taiwanese identity comparing to 0.6, the mean of *Identity* on the left side of the cut-off (the control group).<sup>9</sup>

In column [2], we add the demographic variables to increase the precision of the estimates and lessen the potential bias due to observables' discontinuity. The demographic variables include gender, parents' education level, parents' ethnicity and the *Home-Hoklo-Ratio*. Parents' ethnicity and education level capture the family influence on respondents' national identity. Empirical literature has

<sup>9</sup>The reference mean 0.6 is the simple mean of *Identity* including samples born between Sep 1983 and Aug 1984. For samples born between Mar 1984 and Aug 1984, the mean is 0.65.

shown that parents would invest in strengthening offspring's national identity to be the same as their own identity (Fouka, 2015). In older generation in Taiwan, ethnicity largely determines one's national identity; for example, Mainlanders are much more likely to hold Chinese identity than other ethnic groups.<sup>10</sup>

The magnitude and the significance of the treatment effect in column [2] are similar to column [1]. Column [3] and [4] reports the results with second order polynomials as controls. Estimate of the treatment effect in column [4] is a little bit smaller than the one in column [3]. We note that once taking into account of the size of the standard error of the *TextbookExposure* estimates in column [3] and [4], the estimates are similar statistically speaking. Further, the fact that the second order polynomial term added in column [3] and [4] are not significant and the small differences of the estimates across columns assures us that the estimates are not sensitive to the polynomial specification.

The magnitude of the estimates are also consistent with Figure 3 and Figure 5. We thus confirm that studying *Knowing Taiwan* series do significantly strengthen one's Taiwanese identity, and our confident and conservative measure of the treatment effect is 18 percentage points.

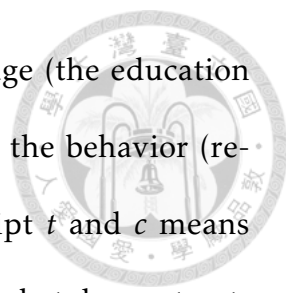
### **Persuasion Rate.**

We provide the persuasion rate calculated by the formula used in media economics (DellaVigna and Gentzkow, 2010):

$$100 \times \frac{y_t - y_c}{e_t - e_c} \times \frac{1}{1 - y_c},$$

---

<sup>10</sup>We define Mainlanders as people with at least one of their parents being Mainlanders; the rest, Others. Combining 2003, 2004, 2005 TSCS waves and keep the samples born between 1950 and 1964, we find that proportion of Mainlanders reporting himself as Taiwanese, Chinese, Both are 25.9%, 10.6%, 61.4% respectively, while the counterpart of Others are 67.2%, 2.7%, 28.7%



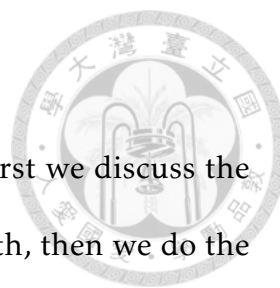
where  $e_i$  denotes the share of the group  $i$  receiving the message (the education content in our case),  $y_i$  denotes the share of group  $i$  adopting the behavior (reporting stronger Taiwanese identity in our case). The subscript  $t$  and  $c$  means treatment and control group. The persuasion rate measures to what degree treatment persuade people into adopting the behavior, scaled by the share of people who receives messages and the space for control group “to be persuaded”  $(1 - y_c)$ . In our case, since all students born after September 1984 are exposed to the new textbook,  $e_t - e_c = 1 - 0 = 1$ . The persuasion rate reported in the first column in Table 2 is calculated by:  $100 \times \frac{0.179}{1} \times \frac{1}{1-0.608} = 45.6$ . This 45% persuasion rate is quite high comparing to the persuasion rate found in media economics literature, which are barely higher than 20%. Our estimates, however, is in align with the persuasion rate found in [Cantoni et al. \(2015\)](#). In their paper, more than a quarter of the persuasion rate is higher than 20%, and the highest one is 41%. The high persuasion rate is not that surprising after taking into account of the degree of the exposure: students have to at least study the *Knowing Taiwan* series for a year and they also have to memorize the materials for the high school admission examinations. The exposure is much greater than specific TV or radio program.



Table 2: Estimated effects of the textbook reform in the Short Run

	Dep Variable: Taiwanese Identity			
	[1]	[2]	[3]	[4]
<i>TextbookExposure</i>	0.179**	0.176**	0.243**	0.213*
	(0.0757)	(0.0770)	(0.0982)	(0.110)
<i>Month</i>	0.000314	0.000225	0.00687	0.00917
	(0.00491)	(0.00503)	(0.0172)	(0.0179)
<i>TextbookExposure</i> <i>×Month</i>	-0.00614	-0.00520	-0.0473**	-0.0422**
	(0.00755)	(0.00748)	(0.0191)	(0.0197)
<i>Month</i> <sup>2</sup>			0.000256	0.000352
			(0.000771)	(0.000798)
<i>TextbookExposure</i> <i>×Month</i> <sup>2</sup>			0.00144	0.00102
			(0.00107)	(0.00107)
Constant	0.603***	0.488***	0.619***	0.517***
	(0.0929)	(0.110)	(0.107)	(0.127)
Control Group Mean	0.608	0.608	0.608	0.608
Persuasion Rate	45.6	44.8	61.9	54.3
Observations	417	417	417	417
R-squared	0.110	0.160	0.118	0.165
Survey Year & Hometown FE	Yes	Yes	Yes	Yes
Demographic	No	Yes	No	Yes

Notes: We pool data from 2003, 2004, and 2005 TSCS waves. We run regression with specification 1. All columns include hometown fixed effect and survey year fixed effect. Column [1] and [2] include first-order polynomials. Column [2] further include demographic control including gender, parents' education level (category variable: elementary school, junior high school, senior high school, senior vocational school, college, university, military), parents ethnicity (category variable: Hoklo, Mainlanders, Hakka, Other), and *Home-Hoklo-Ratio* dummy variable. Column [3] and [4] include second order polynomials. Column [3] does not include demographics while column [4] does. *TextbookExposure* is 1 if the birth month of the respondent is after Sep 1984, 0 otherwise, indicating if one studies the *Knowing Taiwan* series. *Month* is the birth month recentered at Sep 1984. *Month*<sup>2</sup> means square term of *Month*. Standard errors are clustered at birth month level in parentheses \*\*\* p<0.01, \*\* p<0.05, \* p<0.1



## 4.5 Robustness Check

We validate the robustness of the main results in two ways. First we discuss the sensitivity of the results to the choice of birth month bandwidth, then we do the falsification test by changing the birth month cut-off to previous years.

### 4.5.1 Bandwidth Choice

The results in Table 2 are qualitatively similar to a wide range of bandwidths. In Figure 6, we plot the point estimates and the corresponding 95% confidence intervals derived from regression specifying linear polynomials controlling for trend (same specification as in column [1] in Table 2) with bandwidths ranging from two education cohorts to one education cohort on each side of the cut-off.<sup>11</sup> The magnitudes of the point estimate are similar as we narrow down the birth month window, showing that the results in Table 2 are not sensitive to the bandwidth choice.

The change of the size and significance of the estimates appears once we restrict our sample to less than 30 months. Investigating into the graphs presenting *Identity* variation at month level (Figure 20 in Appendix C), we find that the decrease of the magnitude is due to the fact that birth month cohort of  $-15$  and  $-14$  exhibit much more higher *Identity* mean (survey year fixed effects controlled) than later month cohorts. The left-hand side fitted line of *Identity* mean are therefore steeper when we exclude  $-14$  and  $-15$  birth month cohorts, leading to a decrease of the size of the jump implied by the difference of the fitted line around the cut-off. It should be noted that the fitted line of the right hand side

---

<sup>11</sup>We provide the results controlling for demographic variables in Fig 19 in Appendix C

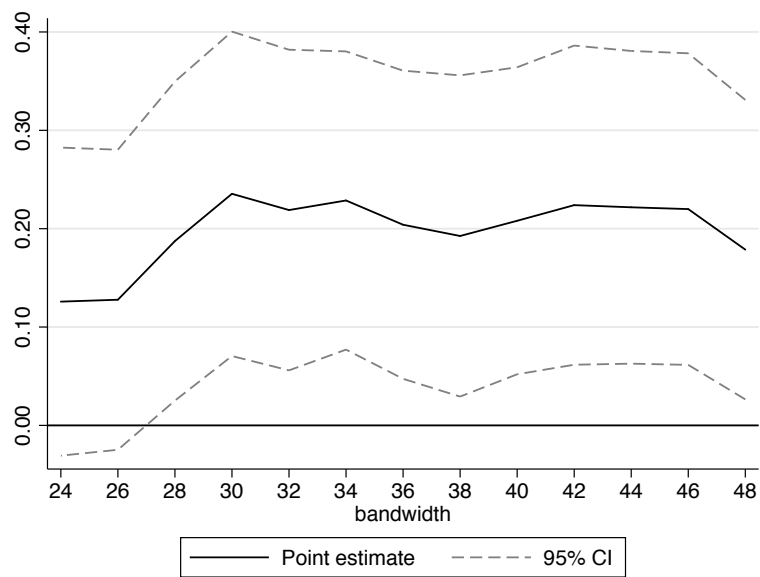


Figure 6: Bandwidth Sensitivity

Notes: We run regressions as column [1] in Table 2 with different bandwidths. The bandwidth 48 means we include 24 birth months on each side of the cut-off, i.e., two education cohorts. The solid line represent the point estimates of the *TextbookExposure* dummy variable and the dotted line represents the corresponding 95% confidence interval derived from standard errors clustered at birth-month level.

(the “treatment group”) barely changes when we confine samples to narrower bandwidths.

The fact that the estimate of the treatment effect decreases when we investigate the behavior within two education cohorts leads to concerns about the overestimate of the size of the treatment effect. We argue that this is not a threat to the interpretation of our main results in Table 2. Observing the 1996 education cohort in Panel C in Figure 20, we could still see that most of the points are below 0 (survey year fixed effects controlled), while most of the dots on the right hand side (the treated group) locate between 0.1 and 0.2; the difference fits our main regression results.



#### 4.5.2 Falsification Test

Another possible explanation of the jump other than the *Knowing Taiwan* reform is mental age effect: people who were born on the exact left hand side of the birth month cut-off are more mentally mature than those on the right hand side since they enter school system earlier and thus, at any given time, they may have more working experience and experience dealing with others (they leave the education system earlier). Being that case, we should observe similar jumps in every yearly cohorts. To examine this speculation, we run regression of equation 1 with moving fake reform windows.

We reckon 1996, 1995 and 1994 as fake textbook reform education year and thus treat September 1983, 1982 and 1981 as fake birth month cut-off. We then replicate the results in Table 2 for each fake textbook reform using the same TSCS waves in 2003, 2004, and 2005. Note that we only include two education cohorts on each side of the fake birth month cut-off to make falsification results comparable to our main results.

Table 3 shows the results of the falsification regressions. Signs of *FakeTextbookExposure* are generally inconsistent across different columns within a fake reform and nearly no significant result is found except the fake reform in 1996 education year. The magnitudes also largely fluctuate across columns within a panel. Furthermore, we plot the quarterly mean of *Identity* (survey year fixed effects controlled) for each fake reform in Appendix D. No obvious discontinuity around the fake birth month cut-off is displayed. Both the regression results and the visual evidence suggest that mental age effect could not explain the jump we see in Figure 5 and the significant treatment effects in Table 2.



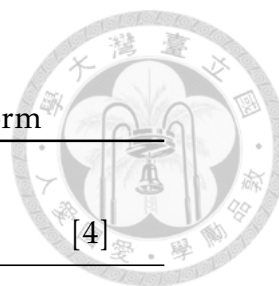


Table 3: Falsification Test of Fake Textbook Reform

	Dep. Variable: <i>Identity</i>			
	[1]	[2]	[3]	[4]
Panel A: September 1983 as Birth Month Cut-Off				
<i>FakeTextbookExposure</i>	-0.0545 (0.0713)	-0.0449 (0.0804)	-0.249** (0.106)	-0.204* (0.117)
Observations	479	479	479	479
R-squared	0.073	0.135	0.083	0.143
Panel B: September 1982 as Birth Month Cut-Off				
<i>FakeTextbookExposure</i>	0.0261 (0.0986)	-0.00500 (0.0994)	0.186 (0.139)	0.187 (0.131)
Observations	500	500	500	500
R-squared	0.067	0.131	0.074	0.140
Panel C: September 1981 as Birth Month Cut-Off				
<i>FakeTextbookExposure</i>	-0.0116 (0.0714)	0.0150 (0.0736)	-0.0201 (0.0950)	-0.0589 (0.0899)
Observations	509	509	509	509
R-squared	0.058	0.172	0.060	0.173
Linear Trend	Yes	No	Yes	No
Quadratic Trend	No	Yes	No	Yes
Survey FE	Yes	Yes	Yes	Yes
Hometown FE	Yes	Yes	Yes	Yes
Demographic	No	Yes	No	Yes

Notes: We pool 2003, 2004, 2005 TSCS data. Panel A includes 1994-1997 education cohorts; Panel B, 1993-1996; Panel C, 1992-1995. We define *FakeTextbookExposure* as respondents born after Sep 1983, 1982 and 1981. Specifications in each column are the same as in corresponding columns in Table 2. Standard errors are clustered at birth month level in parentheses. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1



## 5 Subgroup Analysis

In this section we investigate the heterogeneity of textbook effect along two dimensions: education track and the ethnicity distribution of one's hometown.

We present four piece of evidence in each subgroup analysis, from raw data examination to the regression result of RDD specification, as in section 4.1 and section 4.4. First we look at the proportion of respondents holding stronger Taiwanese identity in each subgroup in different survey years. We then run a simple regression of *Identity* to see if different subgroup exhibit different Taiwanese identity pattern on year cohort basis. The specification is as follows:

$$Identity_{ikjt} = \beta_0 + \beta_1 Z_i + \beta_{2k} Z_i \times \zeta_k + \zeta_k + \eta_j + \delta_t + \epsilon_{ijt}, \quad (2)$$

where  $Z_i$  is a dummy variable indicating the subgroup division (*Edu – Academic* or *Home – Hoklo – Ratio*);  $\zeta_k$  are education cohort dummies;  $\eta_j$  and  $\delta_t$  are hometown and survey year fixed effects. Our interest focuses on  $\beta_{2k}$ . After year cohort analysis, we zoom in and conduct regression discontinuity analysis within each subgroup.

### 5.1 Heterogeneity: Education Track

One dimension to break up the potential treatment effect heterogeneity is to look into the degree of exposure to the new textbooks. Students who pay more efforts on studying the textbooks are associated with higher treatment intensity in the sense that they may memorize more Taiwan-related texts. The ideal proxy of the intensity is the grade of the social subject in the admission examination of senior

high school/senior vocational school.<sup>12</sup> We do not observe this in the TSCS data, but we could distinguish the rough high/low level of efforts students devoted to general academic subjects in junior high school by their choice of the education track.

After compulsory education, students in Taiwan are divided into two educational tracks: the academic and the vocational tracks. The choice of the track is endogenous and would be highly correlated with the efforts student put in studying when in junior high school. For students who had been motivated to pursue more academic knowledge, they would put a great deal of efforts into studying the textbooks so that they could be selected into better senior high schools. On the other hand, common wisdom suggests that parents in Taiwan would encourage students who lack motivation but are adept at obtaining excellent grades (for example, memorize the material more quickly than an average person) to opt for academic instead of vocational track. Consequently, the division of the education track imply the latent exposure to the *Knowing Taiwan* series.

We categorize observations into two groups. Observations with senior high school or university as their final education attainment are labelled with academic track (*Edu – Academic* equals to 1). Other observations, with junior high school, senior vocational school, college and tech university as final education attainment, are labelled with vocational track (*Edu – Academic* equals to 0).

In Figure 7, we plot the simple mean of *Identity* at education cohort level in 2004 and 2005 survey year. The trend of *Identity* before the textbook reform is not parallel, but we observe that in both 2004 and 2005 waves, students in aca-

---

<sup>12</sup>Junior high school graduates, no matter which education track they wish to proceed, took the same national examination and use the scores from that examination to apply for senior high or senior vocational schools

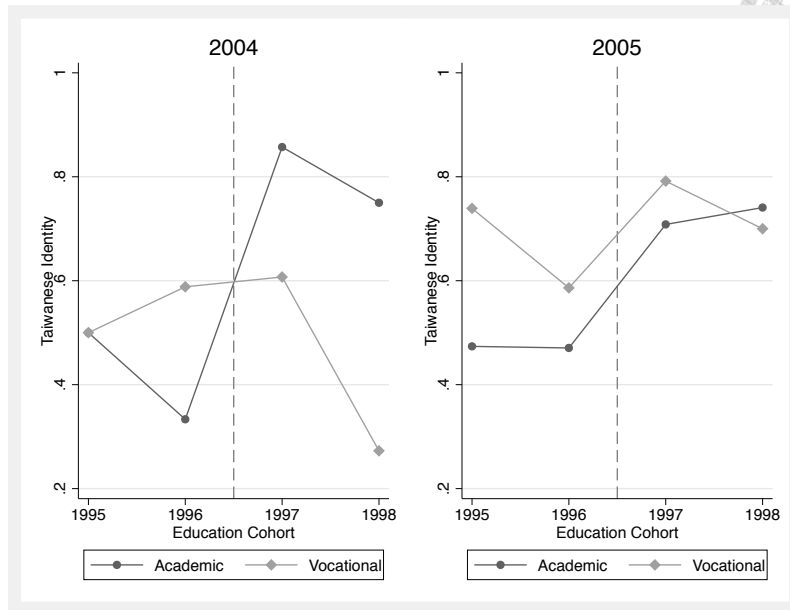


Figure 7: National Identity Trend in Different Education Track

Notes: Data is collapsed at education-cohort-survey year level for observations in two education tracks from 2004 and 2005 waves of TSCS. Dots represent the simple mean of *Identity* at each cell.

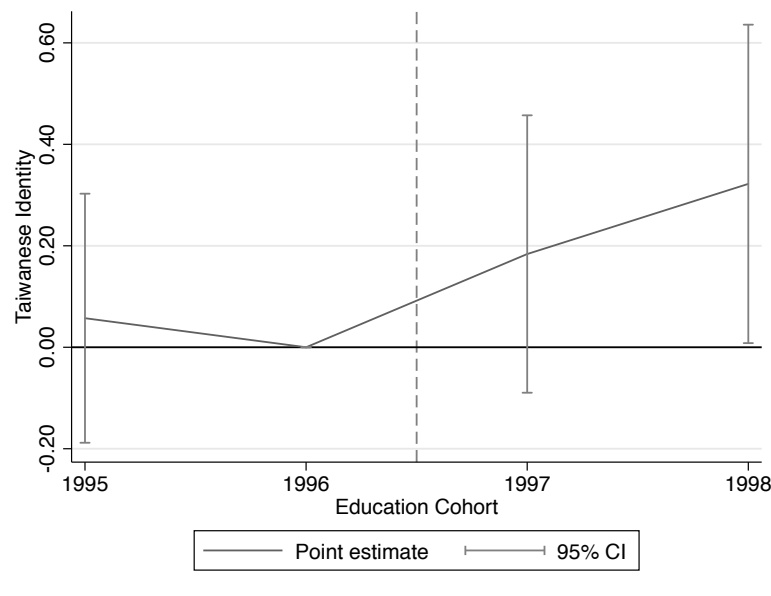



Figure 8: Heterogeneity Textbook Effect in Education Track: Cohort Analysis

Notes: We pool 2003, 2004 and 2005 TSCS waves. Point estimates and confidence intervals of  $\beta_{2k}$  are derived from regression with specification 2 reckoning 1996 education cohort as reference group. Standard errors are clustered at birth month level.



demographic track exhibit more increase of proportion of holding stronger Taiwanese identity once exposed to the textbook reform than students in vocational track. The two dots from the data of 1998 education cohort surveyed in 2004 should be interpreted with caution, since the sample size in each cell is pretty small (4 and 11 for academic and vocational track respectively). We plot the point estimate and confidence interval of regression result of specification 2 in Fig 8. We note that the point estimate of the interaction terms  $\beta_{2k}$  are much more higher for education cohorts exposed to the new textbook than previous ones, though not all significant.

For regression discontinuity analysis, in Figure 9, we plot the mean of *Identity* at birth quarter level (controlling for survey year fixed effect) for academic track (Figure 9a) and vocational track (Figure 9b) observations separately. We observe a manifest jump around the cut-off in Figure 9a and find no systemic pattern of *Identity* mean in Fig 9b. We also note that although no systematic variation is found in Figure 9b, the mean level of the dots before and after the birth-moth cut-off are close, hinting that *Knowing Taiwan* series do not significantly affect the national identity formation for students in vocational track.

Regression results using specification 1 with academic track observations (Panel A) and vocational track observations (Panel B) are provided in Table 4. Results from academic track observations are qualitatively and quantitatively similar across different polynomial settings and the inclusion of demographic variables, while the results from the vocational track are generally not significant and varies in different columns. The estimates show that for academic track students, the new textbook increase their probability of holding stronger Taiwanese identity

by roughly thirty percent, a much larger effect than in the overall samples shown in Table 2 (18%).

One concern over splitting the sample by self education attainment is the endogeneity; namely, the possibility that the introduction of *Knowing Taiwan* series alters the demographic composition of students in two tracks. Specifically, one might suspect that students who originally hold stronger Taiwanese identity may obtain better academic grades (than weaker Taiwanese identity holders), and hence are more likely to enter academic track. We argue that this is implausible. According to the score assignment of the admission exam on different subjects, the materials of *Knowing Taiwan* series only account for one fifteenth of the total scores in the exam.<sup>13</sup>

[Cantoni et al. \(2015\)](#) split observations in a similar way with a different goal. They divide the sample into science and humanity track to tease out the possibility that students holding political ideology similar with the new curricula are more likely to enter Peking University. However, since [Cantoni et al. \(2015\)](#) conduct the survey in Peking University, which consists of students excel at memorizing textbook materials, they could not tell if the new curricula do influence those who does not expend efforts on studying the textbooks. Our results add up evidence in this direction, suggesting that in the case of education content, the degree of exposure matters in terms of the persuasion effect.

---

<sup>13</sup>Another relevant issue is the reform of the admission system for senior high and vocational school. Before 1998 education cohort, students in different areas attend “area union examination” to apply for senior high and vocational schools, whereas for 1998 and later education cohorts, students across Taiwan attend the same examination and admission channels other than application solely based on examination scores are available. We argue that the reform would not pose serious threat to our estimation results since our the main results comes from the variation between 1996 and 1997 education cohorts, who experience the same admission system.

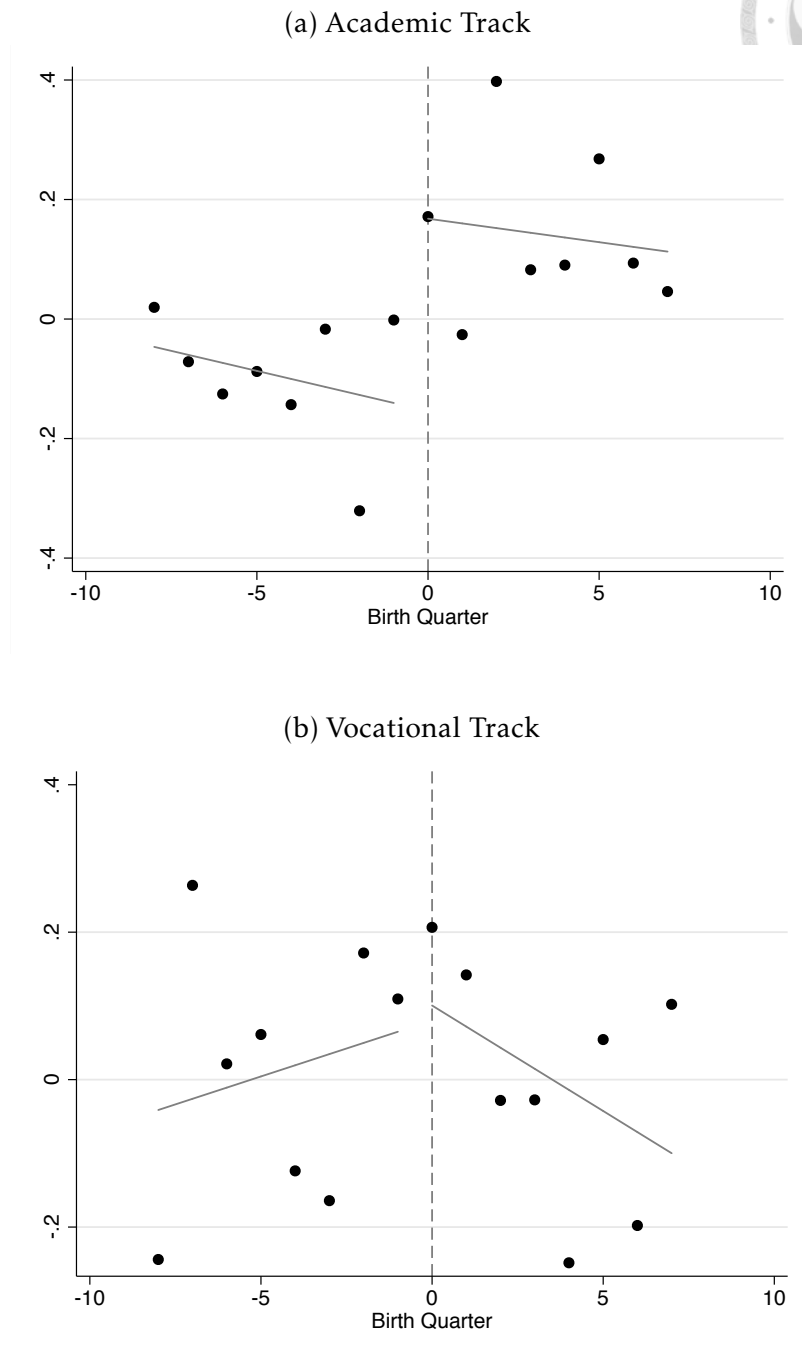


Figure 9: Heterogeneity Textbook Effect in Education Track: RD visualization

Notes: We pool data from 2003, 2004, 2005 TSCS and include education cohorts from 1995 to 1998. Figure 9a includes observations whose final education level is senior high school or university. Figure 9b includes observations whose educational level is junior high school, senior vocational school, college and tech university. We first regress *Identity* on survey year dummies and then collapse the residuals at birth quarter level to derive the dots. Fitted lines are from regression of the dots on a first order polynomial of birth quarter interacted with *KnowingTaiwan* dummy variable.

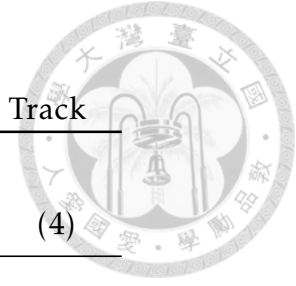


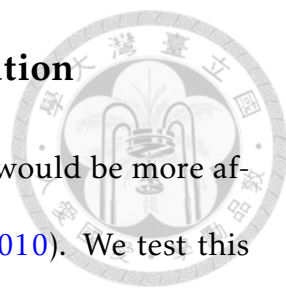
Table 4: Heterogeneity Textbook Effect: Education Track

	Dep. Variable: <i>Identity</i>			
	(1)	(2)	(3)	(4)
Panel A: Academic Track				
<i>KnowingTaiwan</i>	0.303** (0.127)	0.284** (0.127)	0.203 (0.194)	0.231 (0.184)
Control Group Mean	0.580	0.580	0.580	0.580
Persuasion Rate	72.1	66.6	48.3	55
Observations	214	214	214	214
R-squared	0.126	0.213	0.129	0.214
Panel B: Vocational Track				
<i>KnowingTaiwan</i>	0.0474 (0.116)	0.0269 (0.119)	0.226* (0.119)	0.128 (0.139)
Control Group Mean	0.638	0.638	0.638	0.638
Persuasion Rate	13	7.43	62.4	35.3
Observations	203	203	203	203
R-squared	0.183	0.312	0.201	0.320
Hometown FE	Yes	Yes	Yes	Yes
Survey FE	Yes	Yes	Yes	Yes
Linear Trend	Yes	Yes	No	No
Quadratic Trend	No	No	Yes	Yes
Demographic	No	Yes	No	Yes

Notes: We pool 2003, 2004, 2005 TSCS data. Panel A includes observations whose final education level is senior high school or university. Panel B includes observations whose educational level is junior high school, senior vocational school, college and tech university. Specifications in each column are the same as in corresponding columns in Table 2. Standard errors clustered at birth month level in parentheses. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1



## 5.2 Heterogeneity: Hometown Ethnicity Distribution



In “belief-based” models, people who possess less prior belief would be more affected by the new information. (DellaVigna and Gentzkow, 2010). We test this prediction by splitting into two groups: people who lived in towns with more Hoklo people, and otherwise.<sup>14</sup> This division distinguishes the familiarity of students to Taiwan-related knowledge. Children may randomly pick up *cultural parents/role models* in his living neighborhood (Bisin et al., 2011). Since Hoklo people in the older generation typically hold stronger Taiwanese identity, the ratio of Hoklo people surrounding the respondents regulates the probability for a junior high school student picking a role model with strong Taiwanese identity. Furthermore, in terms of political discussions such as election campaigns, people living in towns with less Hoklo people would be exposed to less Taiwan-oriented speeches since politicians running for local elections have to cater to local people’s political preference including national identity.

To show that our criterion distinguishes local environments with different national identity trend. We utilize 1992, 1995 1998 and 2000 waves of TSCS and calculate the mean of *Identity* in *High Hoklo Ratio Towns* and *Low Hoklo Ratio Towns* (divided by the Hoklo people population median in 2004) in different survey years. The results are graphed in Figure 10. It is obvious that the difference between the proportion of people holding stronger Taiwanese identity in these two areas is of size about 0.1 to 0.15. This assures us that students lived in two areas face greatly different society atmosphere in terms of issues regarding na-

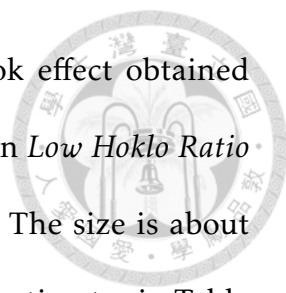
---

<sup>14</sup>We use the variable *Home-Hoklo-Ratio*, which assigns 1 if the proportion of Hoklo people in one’s hometown is less than the population median (77.1%), zero otherwise, to divide the sample. We provide the map demonstrating the Hoklo ethnicity distribution based on *Home-Hoklo-Ratio* in Appendix E.

tional identity when in junior high school and elementary school, the time when they absorb the input for national identity formation from the living environment. We argue that the textbook effect would be greater for students living in *Low Hoklo Ratio Towns* according to “belief-based” models, since they were originally less familiar with Taiwan-related knowledge.

We first look at the raw national identity variation in different survey years. In Fig 11, we observe that in 2005 wave, respondents who lived in *Low Hoklo Ratio towns* are much more likely to hold stronger Taiwanese identity than in 2004 wave, while other respondents who lived in *High Hoklo Ratio towns* do not exhibit such pattern. Despite this change from 2004 to 2005 waves, the jumps of the *Low Hoklo Ratio* line around the cut-off are salient in both survey years and the magnitude of the jump is bigger than the *High Hoklo Ratio* line, which fits our prediction from the belief-based model. We caution that the sample size of the cells of 1998 education cohort in 2004 is small (4 and 11 for *High Hoklo Ratio* and *Low Hoklo Ratio* towns respectively), so one should not make interpretation on the dip of the two dots in 2004. Using regression with specification 2 to test if the differences between the magnitudes of the jumps in two subgroups is significant, we do not, however, find significant results in Figure 12.

To examine the finer variation within an education cohort, in figure 13 we plot the mean of *Identity* at birth quarter level (controlling for survey fixed effect) for observations lived in *Low Hoklo Ratio Towns* (Figure 13a) and *High Hoklo Ratio Towns* (Figure 13b). A clear jump with the rough size 0.35 around the birth month cut-off is exhibited in Figure 13a, while in Figure 13b, the dots does not show meaningful variation.



In Panel A of Table 5, we find that the estimated textbook effect obtained from regression results including only respondents who lived in *Low Hoklo Ratio Towns* with specification 1 is robust across different columns. The size is about 0.35 and significant, consistent with Figure 13 and larger than estimates in Table 2 (overall sample). For respondents who lived in *High Hoklo Ratio Towns*, no significant textbook effect is detected in Panel B of Table 5. Although we observe significant heterogeneity in this dimension, we note that the dummy variable *Home-Hoklo-Ratio* suffers from discontinuity around the birth month cut-off (see Table 1), which might cause potential problem in RD estimations in Table 5.

#### **The Relationship between Main Results and Subgroup Analysis.**

Results of textbook effect along two dimensions, education track and Hoklo people proportion, display significant heterogeneity, The textbook effect interacts with the students' familiarity with the education content, both before (*Home – Hoklo – Ratio* heterogeneity) and after (Education Track heterogeneity).

The finding of textbook effect heterogeneity further helps us rule out two concerns. First, the measurement concern. One might suspect that the main results come from that students in earlier cohorts are afraid of responding their true national identity. If this is the case, then the subgroup analysis would suggest that only students in academic education track and students who lived in *Low Hoklo Ratio Towns* were afraid of responding themselves as Taiwanese, which is very unlikely. For the same reason, if the main results come from other factors such as significant social events, the event should affect students in all subgroups. We are thus more confident that the effect we detect in main results comes from the introduction of the new textbook-*Knowing Taiwan* series.

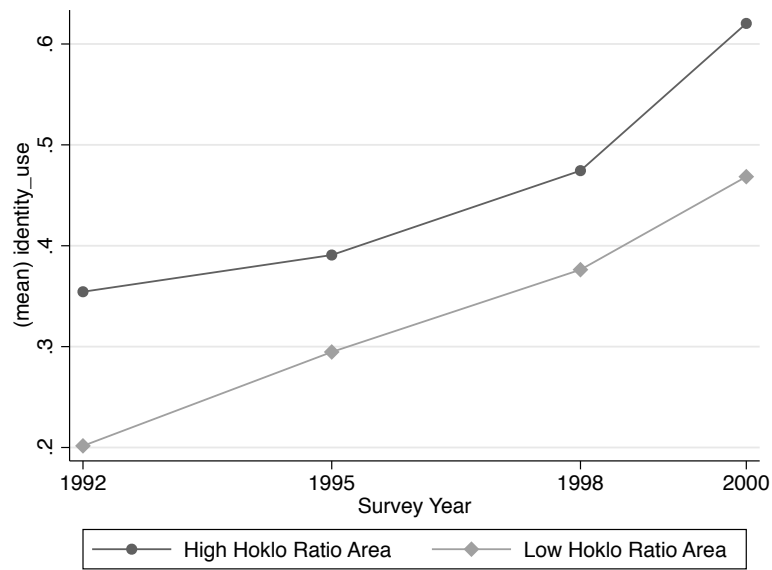
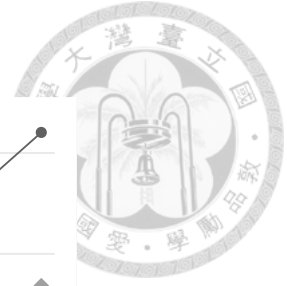


Figure 10: National Identity Trend in High/Low Hoklo Ratio Areas in 1990s

Notes: We pool data from 1992, 1995, 1998 and 2000 TSCS waves. We drop the observations aged below 30 (to include those adults whom children are more likely to meet). We collapse the data into survey-year-Home-Hoklo-Ratio level. Dots represent the mean of *Identity* in each cell.

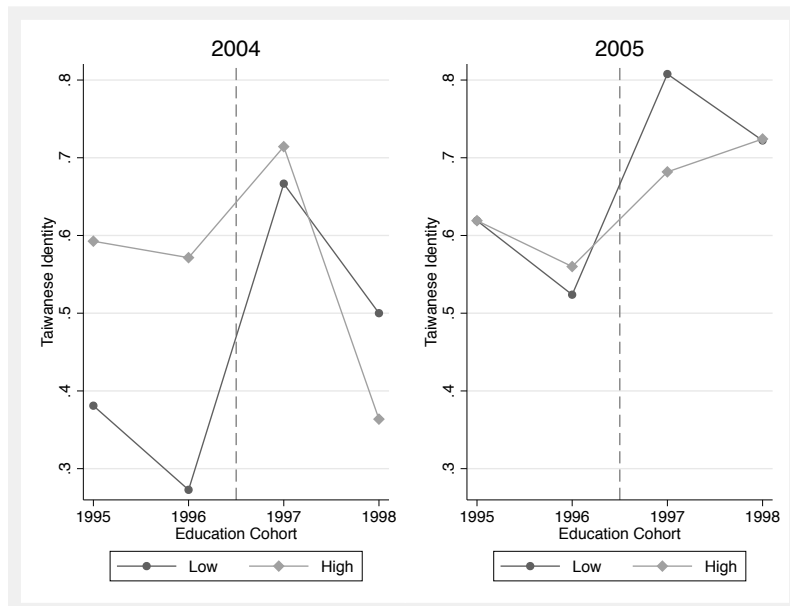


Figure 11: National Identity Trend in High/Low Hoklo Ratio Areas

Notes: Data is collapsed at education-cohort-survey year level for observations in High/Low Hoklo Ratio Areas from 2004 and 2005 waves of TSCS. Dots represent the simple mean of *Identity* at each cell.

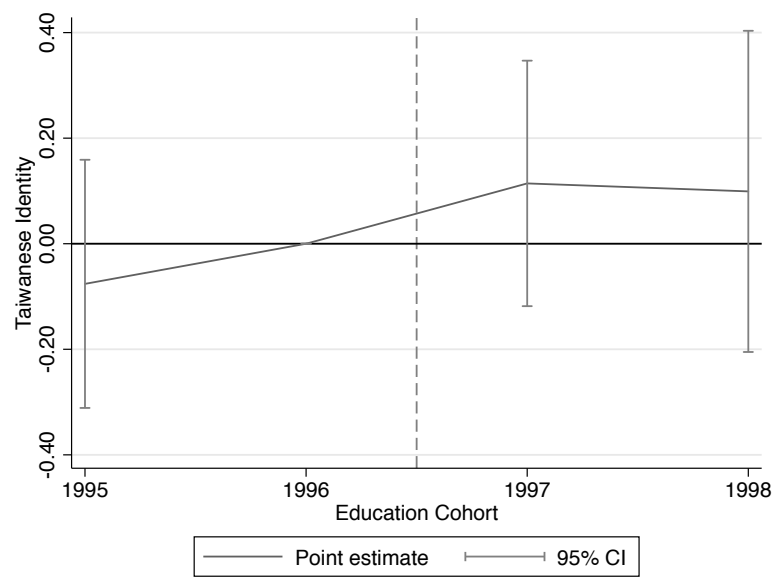


Figure 12: Heterogeneity Textbook Effect in High/Low Hoklo Ratio Areas: Cohort Analysis

Notes: We pool 2003, 2004 and 2005 TSCS waves. Point estimates and confidence intervals of  $\beta_{2k}$  are derived from regression with specification 2 reckoning 1996 education cohort as reference group. Standard errors are clustered at birth month level.

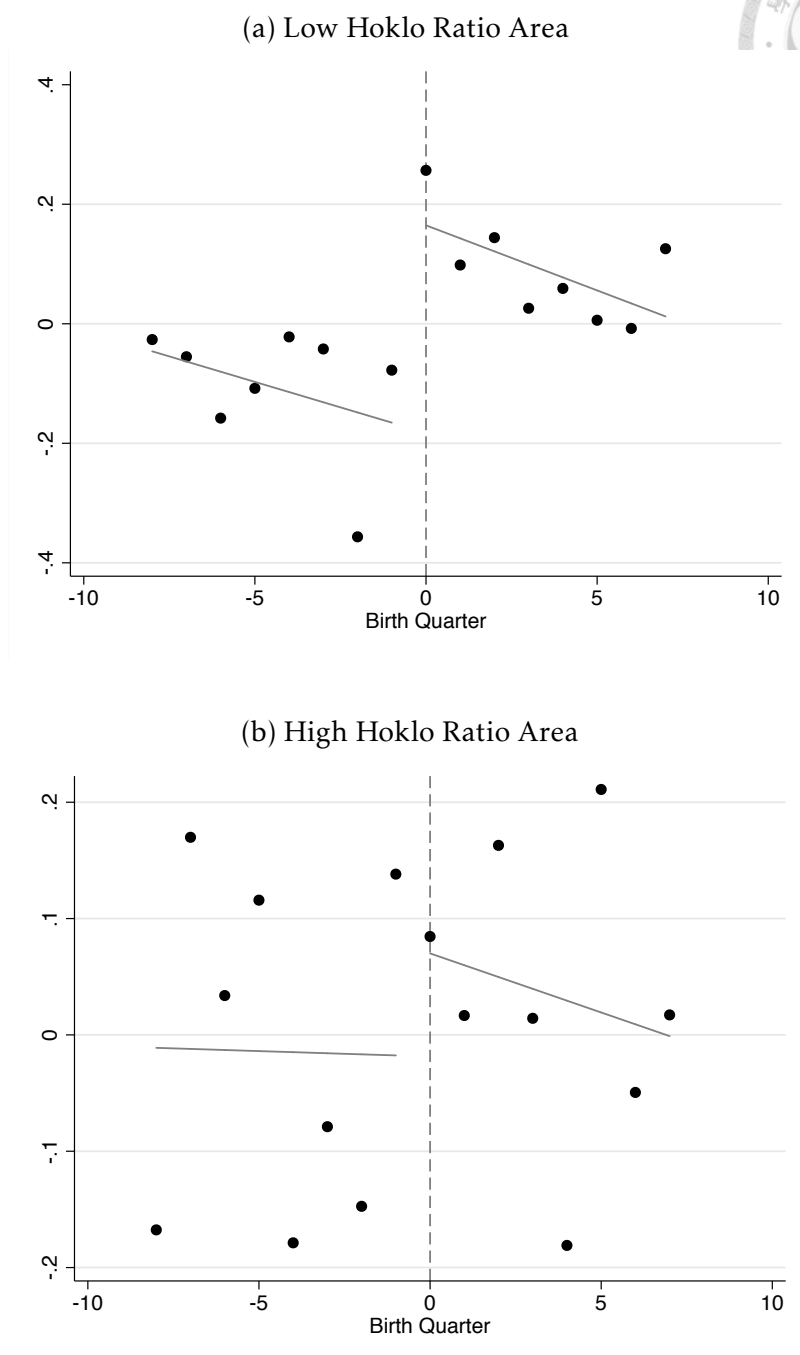


Figure 13: Heterogeneity Textbook Effect in High/Low Hoklo Ratio Areas: RD visualization

Notes: We pool data from 2003, 2004, 2005 TSCS and include education cohorts from 1995 to 1998. Figure 13a includes observations responded lived longest in *Low Hoklo Ratio Towns* (*Home-Hoklo-Ratio* equal to 1) before age 15. Figure 13b includes observations responded lived in *High Hoklo Ratio Towns* (*Home-Hoklo-Ratio* equal to 0) longest before age 15. We first regress *Identity* on survey year dummies and then collapse the residuals at birth quarter level to derive the dots. Fitted lines are from regression of the dots on a first order polynomial of birth quarter interacted with *KnowingTaiwan* dummy variable.

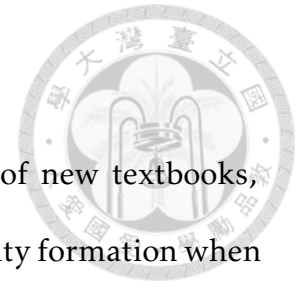


Table 5: Heterogeneity Textbook Effect: Hometown Ethnicity Distribution

	Dep. Variable: <i>Identity</i>			
	(1)	(2)	(3)	(4)
<b>Panel A: Hometown with Low Hoklo People Ratio</b>				
<i>KnowingTaiwan</i>	0.336*** (0.125)	0.366*** (0.125)	0.392** (0.164)	0.423*** (0.155)
Control Group Mean	0.555	0.555	0.555	0.555
Persuasion Rate	75.5	82.2	88	95.7
Observations	186	186	186	186
R-squared	0.147	0.222	0.159	0.230
<b>Panel B: Hometown with High Hoklo People Ratio</b>				
<i>KnowingTaiwan</i>	0.0311 (0.0976)	0.0321 (0.0980)	0.0970 (0.133)	0.0790 (0.148)
Control Group Mean	0.647	0.647	0.647	0.647
Persuasion Rate	8.81	9.09	27.4	22.3
Observations	231	231	231	231
R-squared	0.137	0.213	0.141	0.215
Hometown FE	Yes	Yes	Yes	Yes
Survey FE	Yes	Yes	Yes	Yes
Linear Trend	Yes	Yes	No	No
Quadratic Trend	No	No	Yes	Yes
Demographic	No	Yes	No	Yes

*Notes:* We pool 2003, 2004, 2005 TSCS data. Panel A includes observations whose hometown has lower Hoklo people ratio comparing to the median of the population in National Hakka Population Basic Information Survey Research, while Panel B includes observations whose hometown has higher ratio. Specifications in each column are the same as in corresponding columns in Table 2. Standard errors clustered at birth month level in parentheses. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1

## 6 Long Run Results

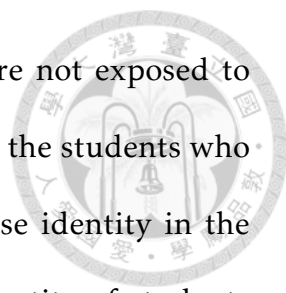


We have until this section established that the introduction of new textbooks, *Knowing Taiwan* series, has impacts on students' national identity formation when they were 18 to 23 years old (short run sample). The natural question emerges: whether the impact is transitory or persistent. Clots-Figueras and Masella (2013) measure the Catalan identity and observe the voting behavior when samples are roughly 25 to 30 years old. Fouka (2015) used ethnic intermarriage and off-spring first names to measure the strength of one's ethnic identity. Both paper found significant language policy effect in such long run outcomes. We explore this issue by examining the long run sample (surveyed from 2010 to 2015, respondents age 24 to 33).

We first look at the education cohort variation in Figure 14. Surprisingly, although the 1998 education cohort exhibits stronger Taiwanese identity, the 1997 education cohort is not significantly more likely to identify himself as Taiwanese. We again examine the finer variation within education cohorts. In Figure 15 we plot the *Identity* mean at birth quarter level (controlling survey year fixed effects). No significant jump is found, and the mean level of *Identity* is about the same on either side of the cut-off. The regression results in Table 6 fit the graph and shows no sign of textbook effect in the long run. We note that the size of the coefficients of *TextbookExposure* are all small across specifications, compared to the results derived from short run sample.

This result, whether the students were exposed to the new textbooks cannot predict the long run national identity difference, prompts us to ask that *which* group's national identity change in the long run, the control, the treatment, or





both. We aim to see that whether it is the students who were not exposed to the textbook are more likely to hold Taiwanese identity, or it is the students who were exposed to the textbook are less likely to hold Taiwanese identity in the long run. Figure 16 offers an interesting answer: national identity of students not exposed to the new textbook *catch up* with the students who were exposed to the new textbook.<sup>15</sup> In Figure 16, we plot the simple mean of the *Identity* dummy variable in treatment group (people born in 1997 and 1998 education cohort) and control group (people born in 1995 and 1996 education cohort). We find that the proportion of people holding stronger Taiwanese identity in control group increases obviously from 2009 to 2015, when the respondents are around their thirties.

One hypothesis of why the national identity in both groups converge in the long run is the peer interaction. Although students who studied the old textbook were originally endowed with weaker Taiwanese identity comparing to students exposed to the new textbook (in the short run), the “old textbook” effect may be lessened by the peer interaction because the former student encounter more and more “Taiwanese-oriented” peers surrounding their living environment. Another hypothesis is that the control group is affected by the overall national identity trend illustrated in Figure 1: an increasing trend of Taiwanese identity. To explore if any one of these two are more aligned with the data is a potential research direction.

---

<sup>15</sup>We do not use the data from 2003 in the following analysis since the sample size of treatment group in 2003 is too small. We include the 2009, 2015 data in the following analysis. For 2009 data, it is because we now do not require the parents’ education variables. For 2015 data, I will incorporate it in other parts of analysis in the future.

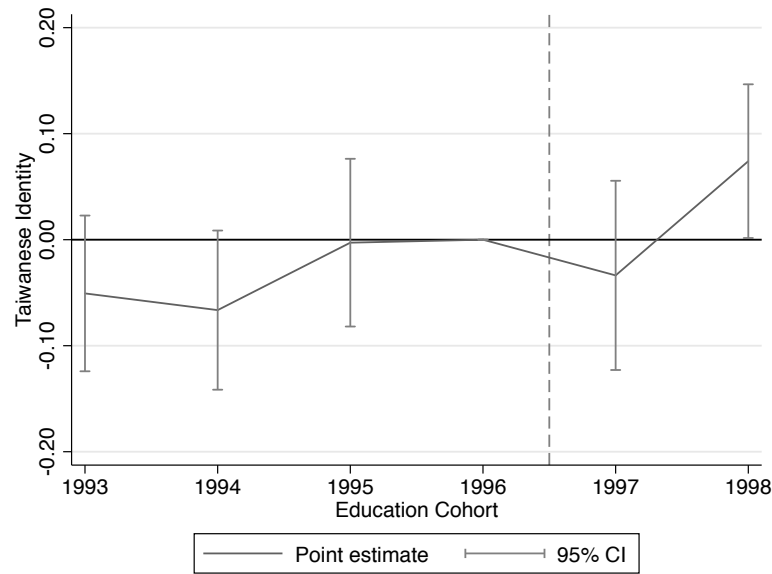
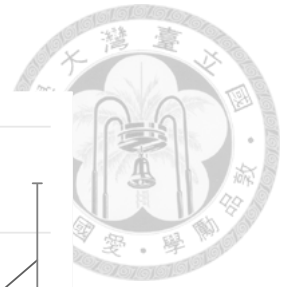


Figure 14: Estimated Textbook Effect by Education Cohort in the Long Run

Notes: We pool 2010, 2012, 2013, 2014 TSCS waves. Point estimates derived from regression of *Identity* on survey year dummies and education cohort dummies (1996 as reference group). Standard errors are clustered at birth month level.

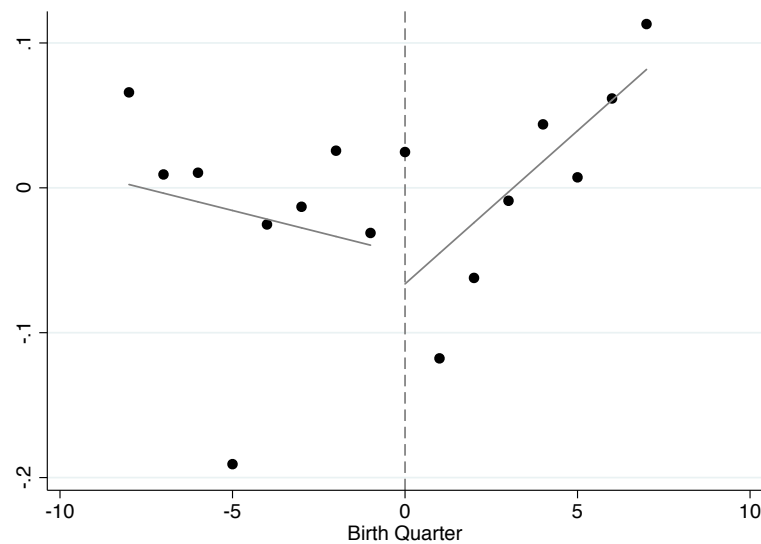


Figure 15: Long Run National Identity Variation at Birth Quarter Level

Notes: I pool data from 2010, 2012, 2013, 2014 and 2015 TSCS waves and include education cohort from 1995 to 1998. We first regress *Identity* on survey year dummies and then collapse the residuals at birth quarter level to derive the dots. Fitted lines are from regression of the dots on a first order polynomial of birth quarter interacted with *KnowingTaiwan* dummy variable.

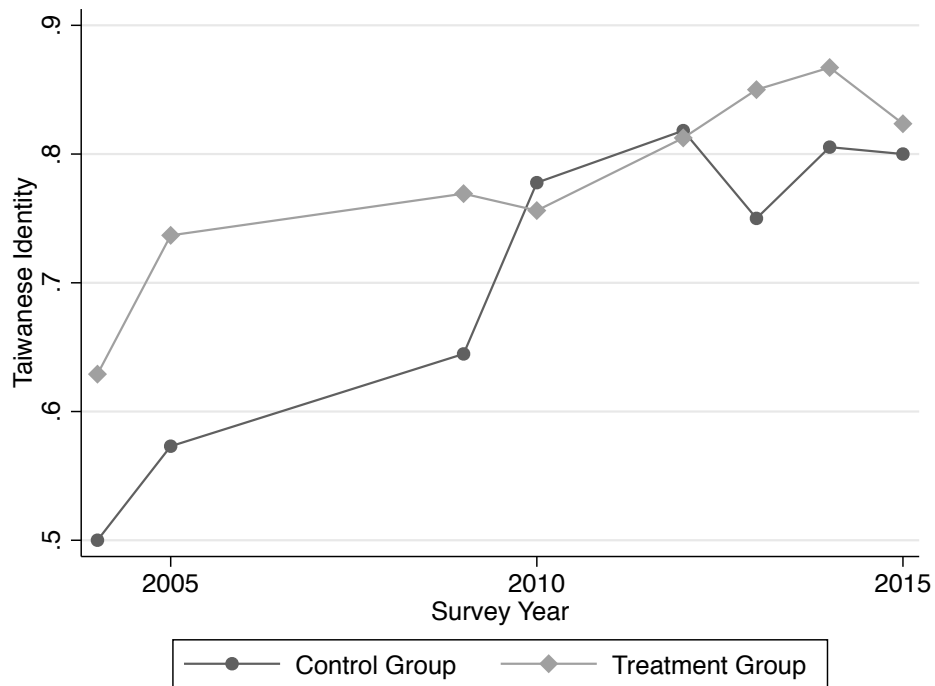


Figure 16: National Identity Variation by Treatment Status

Notes: I pool data from 2004, 2005, 2010, 2012, 2013, 2014 and 2015 TSCS and include education cohort from 1995 to 1998. Dots represent simple mean of *Identity* of each cell in different survey years. Control group includes 1995 and 1996 education cohorts and Treatment group includes 1997 and 1998 education cohorts

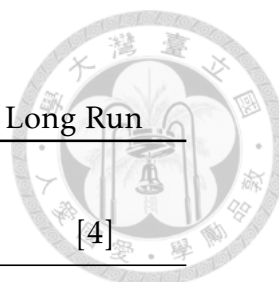


Table 6: Estimated effects of the textbook reform in the Long Run

	Dep. Variable: <i>Identity</i>			
	[1]	[2]	[3]	[4]
<i>TextbookExposure</i>	-0.0256 (0.0667)	-0.0313 (0.0692)	-0.0510 (0.0992)	-0.0661 (0.107)
<i>Month</i>	-0.00240 (0.00205)	-0.000606 (0.00251)	0.0129 (0.0133)	0.0153 (0.0149)
<i>TextbookExposure</i> $\times$ <i>Month</i>	0.00958** (0.00443)	0.00614 (0.00476)	-0.0172 (0.0198)	-0.0194 (0.0215)
<i>Month</i> <sup>2</sup>			0.000605 (0.000506)	0.000629 (0.000543)
<i>TextbookExposure</i> $\times$ <i>Month</i> <sup>2</sup>			-0.000115 (0.000794)	-0.000219 (0.000781)
Constant	0.964*** (0.0781)	1.033*** (0.124)	1.038*** (0.0859)	1.110*** (0.146)
Control Mean	0.78	0.78	0.78	0.78
Observations	535	535	535	535
R-squared	0.057	0.161	0.061	0.164
Hometown FE	Yes	Yes	Yes	Yes
Survey FE	Yes	Yes	Yes	Yes
Demographic	No	Yes	No	Yes

Notes: We pool the 2010, 2012, 2013, 2014 and 2015 TSCS waves and include two education cohorts on both sides of the birth-month cut-off (Sep 1984). Specifications are the same as Table 2. Clustered standard errors at birth month level in parentheses.\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

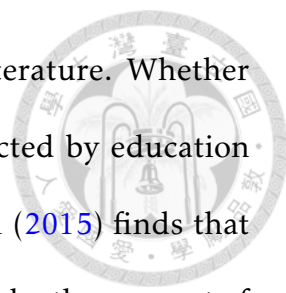
## 7 Conclusion



Can education content influence national identity formation? Although the effect of different education policy (e.g., language use and curriculum reform) on various preferences has been found in recent literature ([Clots-Figueras and Masella, 2013](#); [Fouka, 2015](#); [Cantoni et al., 2015](#)). Direct relationship between the two has not been established yet. We consider the textbook reform which introduced new textbooks containing much more Taiwan-oriented materials (*Knowing Taiwan* series). We adopt education cohort comparison and regression discontinuity design (RDD) on birth month level to tease out the national identity society trend. We find large and significant effect of the education content on holding stronger Taiwanese identity.

We interpret the result coming from the priming effect of “Taiwan” and the instillation of Taiwan-related knowledge. Our results suggest that education content could not only influence the preference over political institutions ([Cantoni et al., 2015](#)), but also national identity, providing more empirical evidence of the effect of education on homogenizing people, as advocated in theory literature ([Alesina and Reich, 2015](#)). Nonetheless, since we measure the national identity by self response of survey questions instead of real behavior, the next step would be examining the effect of education content upon behavior such as voting.

We contribute to the literature by presenting three facts. First, the degree of exposure of education content matters. Students who studied the new textbook harder, measured by the academic/vocational track division, are more affected by the new textbook. Second, the education content effect is larger for students holding weaker prior beliefs (identified by the hometown ethnicity dis-



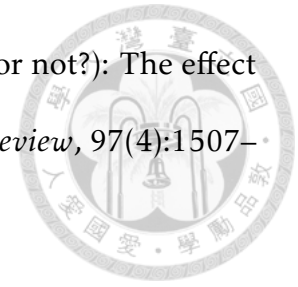
tribution), align with the belief based models in persuasion literature. Whether people holding stronger or weaker prior beliefs are more affected by education policies are not certain in the literature. [Voigtländer and Voth \(2015\)](#) finds that people who held stronger prior anti-Semitic attitude (identified by the support of anti-Semitic parties in previous election) are more affected by the anti-Semitic indoctrination between 1933 and 1945; they exhibit largest increases in anti-Jewish attitudes. Why the persuasion is effective in different subgroups under different contexts is a potential research question in the future. The two findings, the importance of degree of exposure and the prior belief, generalize the findings in [Cantoni et al. \(2015\)](#), which documents the textbook effect in Peking university students, a group of students potentially more affected by the textbooks due to their hard work in senior high school.

Finally, we document that the textbook effect is *not* permanent by showing that whether the students study the new textbook could not predict significant national identity differences in the long run. We propose peer interaction channel and exposure to society trend as explanation but more examination is required. Based on our finding, the long run effect of the curriculum reform in China studied by [Cantoni et al. \(2015\)](#), which changes students ideology significantly when they are still enrolled in Peking University, is thus an exciting topic for future research.

## References

Aghion, P., Jaravel, X., Persson, T., and Rouzet, D. (2014). Education and military rivalry. *Working Paper*.

Alesina, A. and Fuchs-Schundeln, N. (2007). Good-bye lenin (or not?): The effect of communism on people's preference. *American Economic Review*, 97(4):1507–1528.



Alesina, A. and Reich, B. (2015). Nation building. *Working Paper*.

Bisin, A., Patacchini, E., Verdier, T., and Zenou, Y. (2011). Formation and persistence of oppositional identities. *European Economic Review*, 55(8):1046–1071.

Cantoni, D., Chen, Y., Yang, D. Y., Yuchtman, N., and Zhang, Y. J. (2015). Curriculum and ideology. (*forthcoming*) *Journal of Political Economy*.

Clots-Figueras, I. and Masella, P. (2013). Education, language, and identity. *The Economic Journal*, 123:F332–F357.

DellaVigna, S. and Gentzkow, M. (2010). Persuasion: Empirical evidence. *Annual Review of Economics*, pages 643–669.

DellaVigna, S. and Kaplan, E. (2007). The fox news effect: Media bias and voting. *The Quarterly Journal of Economics*, 122(3):1187–1234.

Fan, E., Liu, J.-T., and Chen, Y.-C. (2014). Is the 'quarter of birth' endogeneous? evidence from one million sibilings in taiwan. *NBER Working Paper Series*.

Fouka, V. (2015). Backlash: The unintended effects of language prohibition in us schools after world war i. *Working Paper*.

Liu, M.-H., Huang, L.-C., and Vickers, E. (2005). *History Education and National Identity in East Asia*, chapter Identity Issues in Taiwan's History Curriculum. Routledge Taylor Francis Group, 1 edition.

Shigeoka, H. (2014). The effect of patient cost sharing on utilization, health, and risk protection. *American Economic Review*, 104(7):2152–2184.

Voigtländer, N. and Voth, H.-J. (2015). Nazi indoctrination and anti-semitic beliefs in germany. *Proceedings of the National Academy of Sciences of the United States of America*, 112(26):7931–7936.

Wang, F.-C. (2001). National imagination, ethnic consciousness, and history: Content and context analysis of the “getting to know taiwan” textbook disputes,. *Taiwan Historical Research*, 8(2):145–208.



# Appendices



## A National Identity Distribution and Sample Selection

Table 7: National Identity Distribution in Main Regression Sample

Identity	Taiwanese	Both (Taiwanese and Chinese)	Chinese	Total
#Obs	269	132	16	417
Percent	64.51	3.84	31.65	100

*Notes:*

We pool 2003 2004 2005 TSCS waves and include two education cohorts on both sides of the birth-month cut-off (Sep 1984). The identity is measured by the response of two questions:

1. "In our society, somebody call themselves "Taiwanese," somebody call themselves "Chinese," and somebody call themselves "both." Do you consider yourself as "Taiwanese," "Chinese," or "Both"?"
2. "Below are some possible descriptions of yourself. Which is most suitable to you?"

We define Taiwanese as respondents answering Taiwanese, Chinese to those answering Chinese, and Both to those answering Both, Taiwanese and Chinese, Chinese and Taiwanese.

Table 8: Samples Dropped by Sample Selection

Rule	Short Run Sample	Long Run Sample
	#Obs	#Obs
Reports born outside Taiwan or lived longest outside Taiwan before 15	0	10
Lack of Demographic Info	11	45
Answer Other in identity measure question	8	14
Total drop ratio	4.5%	12.8%
Final sample Size	417	535

*Notes:* The short run sample pools data from 2003, 2004, 2005 TSCS waves and the long run sample pools 2010, 2012, 2013, 2014 waves. We include two education cohorts on each side of birth month cut-off.

## B Observable Characteristics

Table 9: Observables Continuity with Polynomials of Second Order

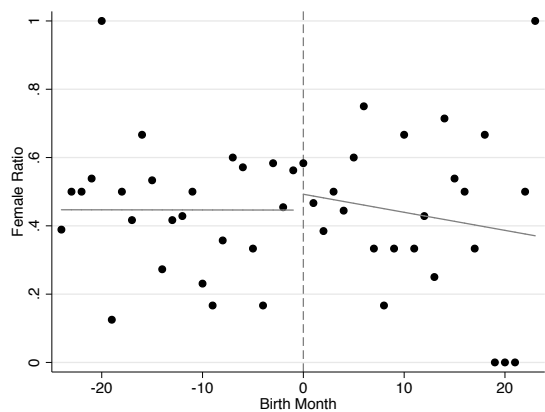
VARIABLES	gender	Father Ethnicity	Mother Ethnicity	Father Edu Level	Mother Edu Level	Father Edu Years	Mother Edu Years	Self Edu Track	Self Edu Years	Hometown Hoklo Ratio
<i>TextbookExposure</i>	-0.0188 (0.0847)	-0.0975 (0.101)	-0.0172 (0.0766)	-0.366*** (0.134)	-0.125 (0.123)	-1.589 (1.050)	-0.648 (0.676)	0.0358 (0.0862)	-0.578* (0.310)	0.256** (0.105)
<i>Month</i>	0.0170 (0.0132)	0.0108 (0.0135)	-0.00198 (0.0142)	0.0387** (0.0174)	0.0217 (0.0130)	0.238 (0.152)	0.156* (0.0894)	0.00890 (0.0119)	0.0521 (0.0493)	-0.0247 (0.0164)
<i>Month</i> <sup>2</sup>	0.000643 (0.000512)	0.000339 (0.000645)	-8.37e-05 (0.000645)	0.00149** (0.000577)	0.000884* (0.000455)	0.00841 (0.00509)	0.00531 (0.00327)	0.000228 (0.000501)	0.00316 (0.00205)	-0.000815 (0.000566)
<i>TextbookExposure</i> <i>xMonth</i>	-0.0266 (0.0219)	0.00653 (0.0219)	0.0187 (0.0197)	0.0169 (0.0237)	-0.0185 (0.0240)	-0.00608 (0.182)	-0.135 (0.151)	-0.0468** (0.0219)	0.0216 (0.0929)	0.0100 (0.0221)
<i>TextbookExposure</i> <i>xMonth</i> <sup>2</sup>	-0.000451 (0.00107)	-0.00144 (0.000945)	-0.000966 (0.000865)	-0.00386*** (0.000900)	-0.000947 (0.00101)	-0.0198*** (0.00662)	-0.00626 (0.00647)	0.00144 (0.00102)	-0.00630 (0.00439)	0.00121 (0.000883)
Constant	0.528*** (0.0697)	0.297*** (0.0523)	0.165*** (0.0588)	0.703*** (0.114)	0.548*** (0.0766)	11.67*** (0.999)	10.53*** (0.481)	0.589*** (0.0504)	14.24*** (0.225)	0.291*** (0.0971)
Observations	417	417	414	417	417	417	417	417	417	417
R-squared	0.005	0.009	0.008	0.031	0.005	0.018	0.007	0.011	0.017	0.011

Notes: We pool data from 2003, 2004, and 2005 TSCS waves. We run regression with specification 1. Second order polynomial is included and we do not include demographics. *TextbookExposure* is 1 if the birth month of the respondent is after Sep 1984, 0 otherwise, indicating if one studies the *Knowing Taiwan* series. *Month* is the birth month recentered at Sep 1984. The construction of the variables: Gender: female being 1, male being 0. Father/Mother ethnicity: Hoklo fathers/mothers being 0, otherwise 1. Father /Mother education level: father/mother with no education, elementary school, and junior high school education level being 0. Father/Mother/Self education years is discrete variable range from 0 to 16. Self education track: respondents with final education attainment as senior high school, university being 1, otherwise, 0. Hometown Hoklo Ratio is the dummy variable *Home-Hoklo-Ratio*, construction see section 3 **The proxy of hometown ethnicity distribution**. Clustered standard error at birth month level in parentheses.\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

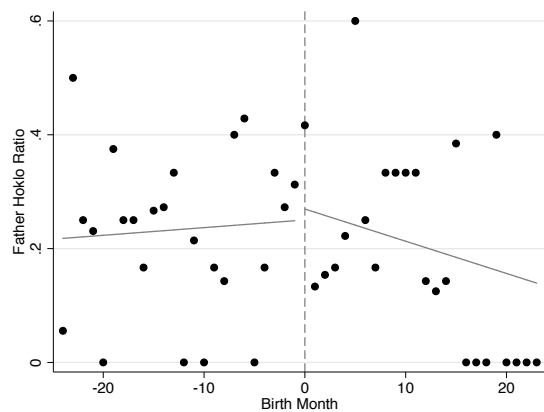




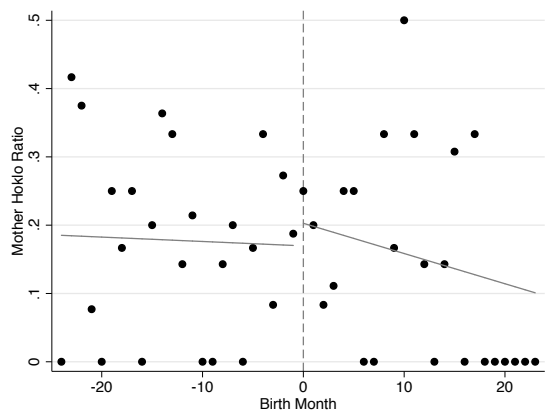
(a) Female Ratio



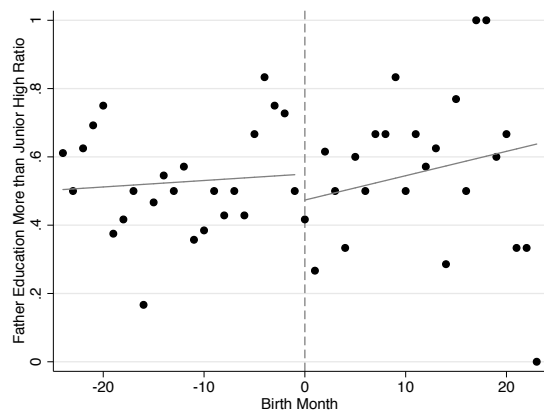
(b) Father Hoklo Ethnicity Ratio



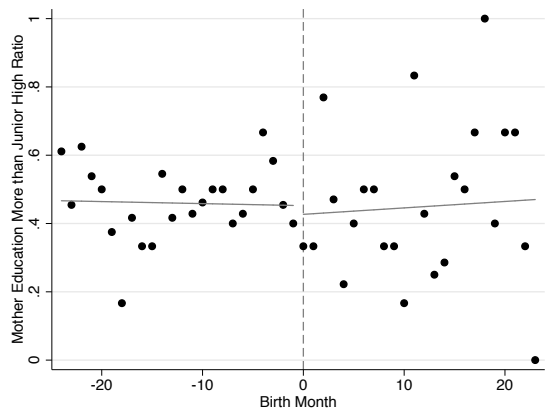
(c) Mother Hoklo Ethnicity Ratio



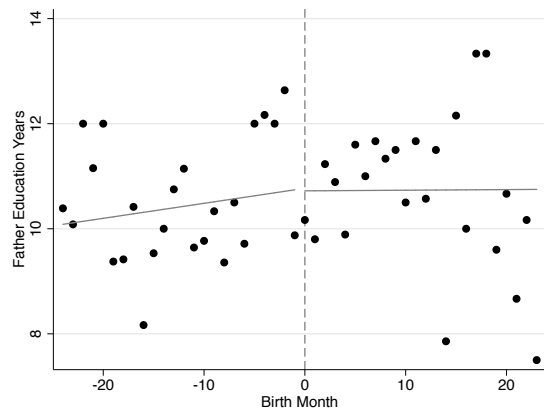
(d) Father Education Level



(e) Mother Education Level



(f) Father Education Years



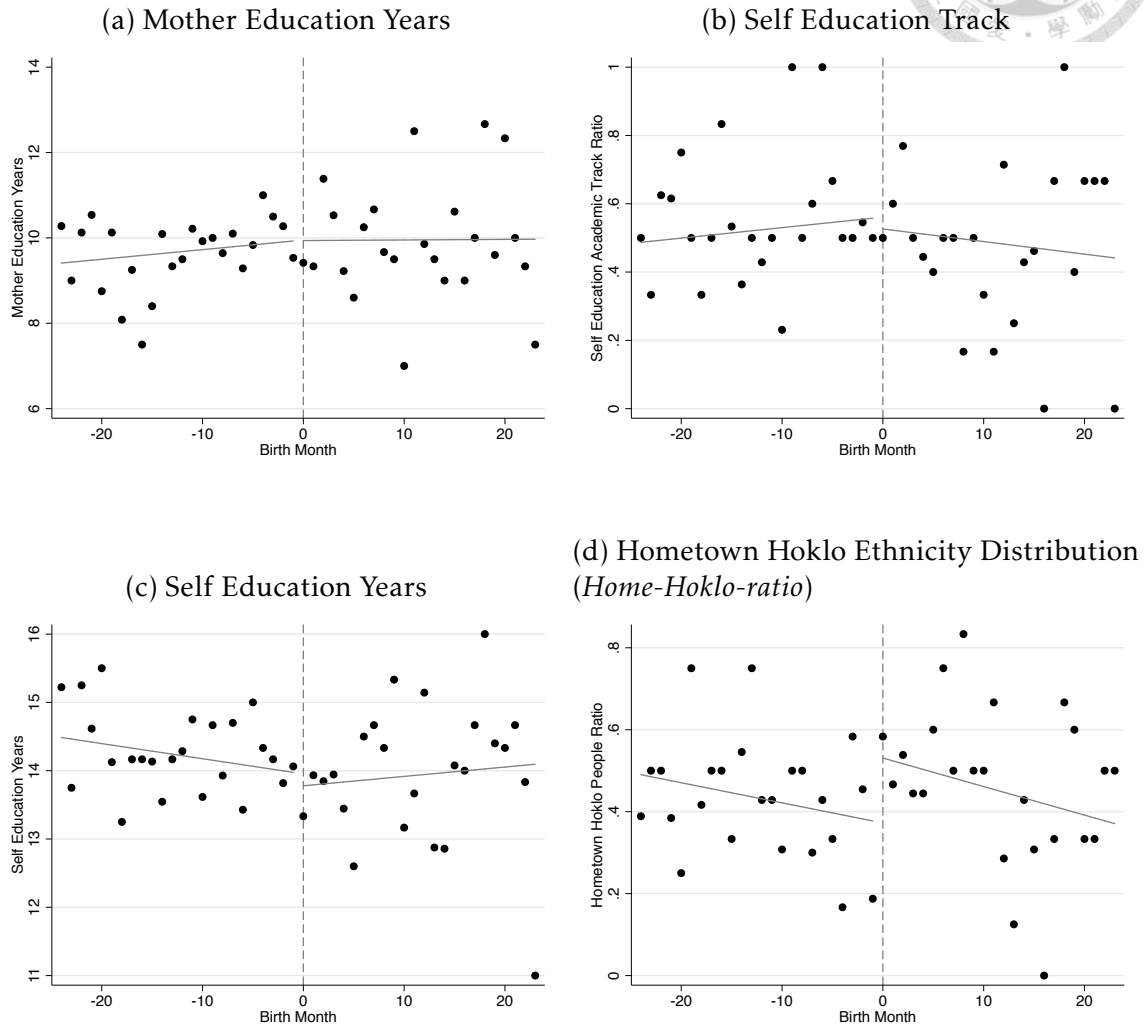


Figure 18: Observable Discontinuity

*Notes:* We pool 2003 2004 2005 TSCS waves and include education cohorts from 1995 to 1998. We first regress the corresponding demographic variable on survey year dummies and then collapse the residuals at birth month level to derive the dots. Fitted lines are from regression of the dots on a first order polynomial of birth month interacted with *TextbookExposure* dummy variable. The construction of the independent variables is the following: Gender: female being 1, male being 0. Father/Mother ethnicity: Hoklo fathers/mothers being 0, otherwise 1. Father/Mother education level: father/mother with no education, elementary school, and junior high school education level being 0. Father/Mother/Self education years is discrete variable range from 0 to 16. Self education track: respondents with final education attainment as senior high school, university being 1, otherwise, 0. Hometown Hoklo Ratio is the dummy variable *Home-Hoklo-Ratio*, see construction in section 3, **The proxy of hometown ethnicity distribution**.

## C Bandwidth Sensitivity

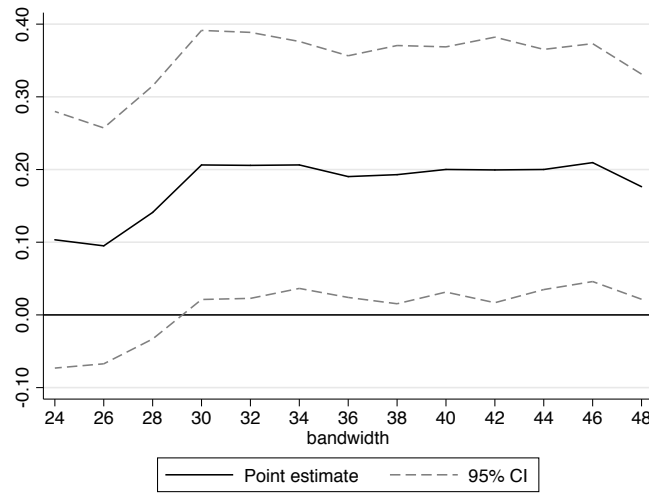


Figure 19: Bandwidth Sensitivity with Demographic Controls

*Notes:* We run regressions as column [2] in Table 2 with different bandwidths. The bandwidth 48 means we include 24 birth months on each side of the cut-off, i.e., two education cohorts. The solid line represent the point estimate of the *TextbookExposure* dummy variable and the dotted line represents the corresponding 95% confidence interval derived from standard errors clustered at birth-month level.

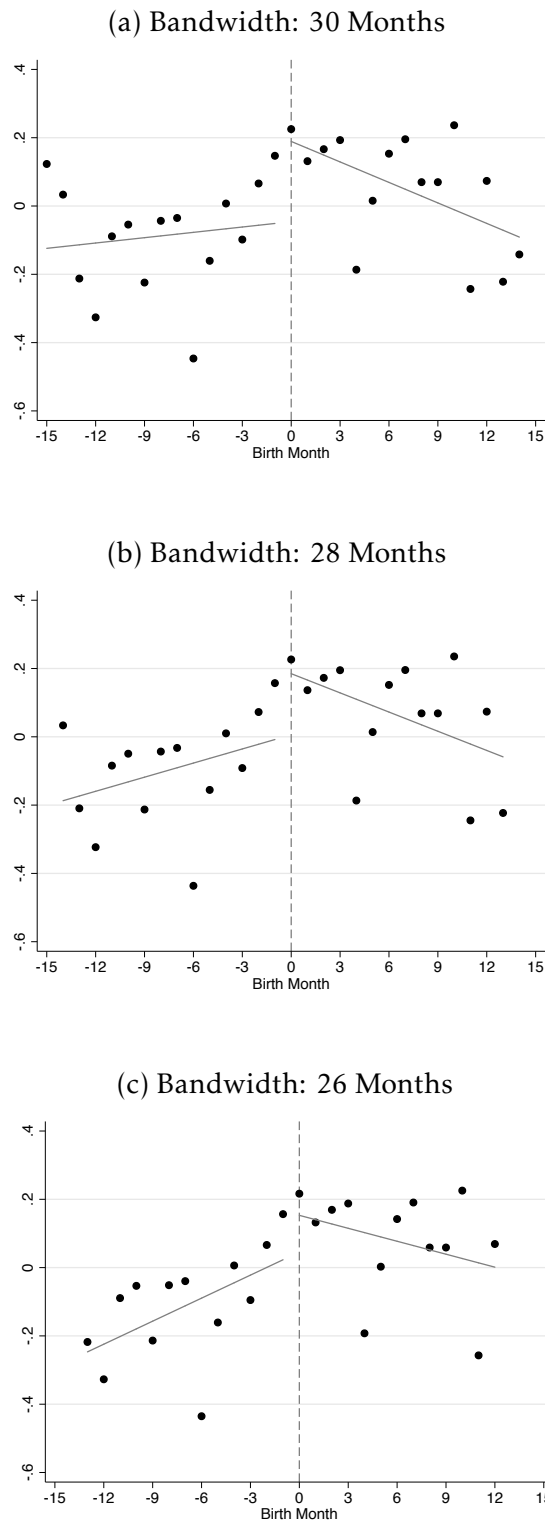


Figure 20: National Identity Variation at Birth Month Level with Different Bandwidths

Notes: We collapse the residuals at birth month level after regressing *Identity* on survey year fixed effects. Each marker represent the simple mean of the residuals in each birth month. Fitted lines are derived from regression of the dots using linear polynomials of birth month interacted with the dummy variable *TextbookExposure*. We confine samples into 15, 14, 13 birth months on each side of the cut-off respectively in Figure 20a, 20b, 20c

## D Falsification Test

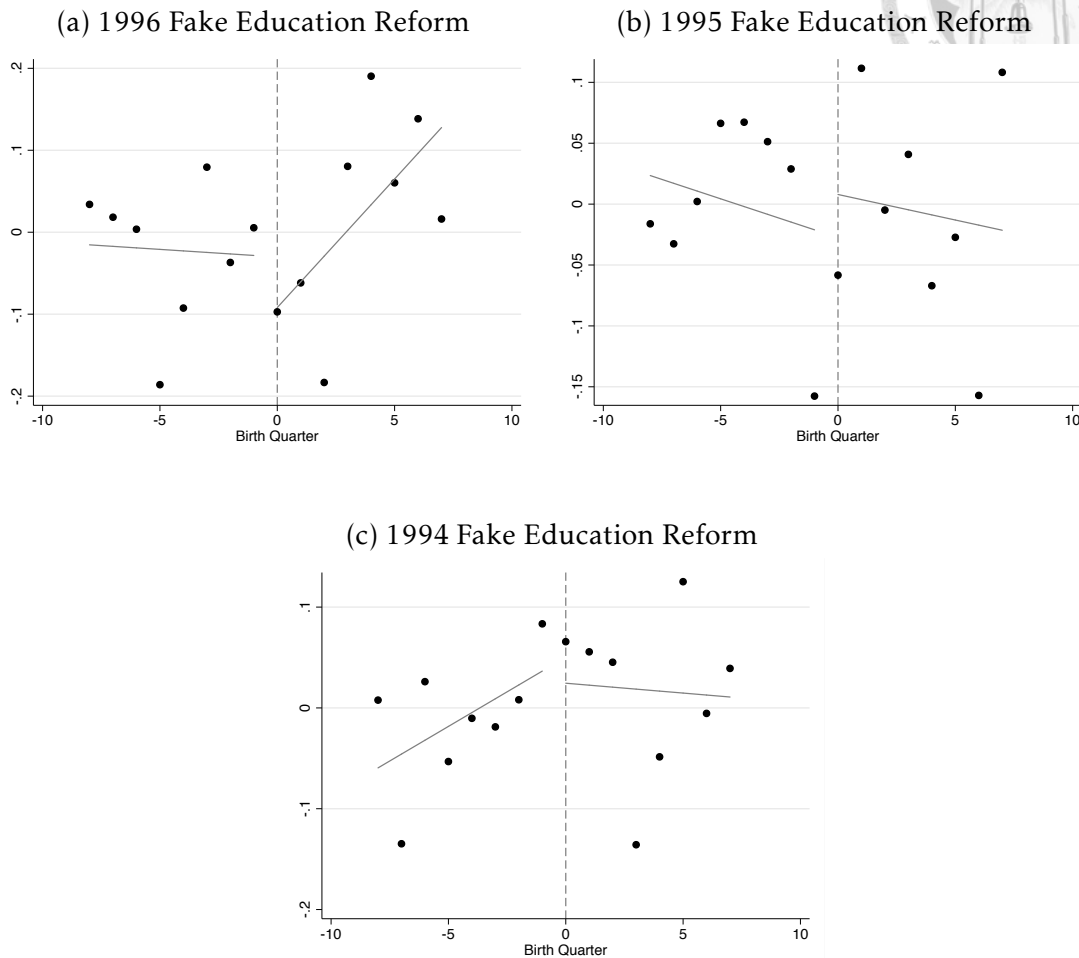


Figure 21: National Identity Variation with **Fake** Birth Month Cut-Off

*Notes:* We pool 2003, 2004, 2005 TSCS waves and generate fake birth month cut-off in Sep 1983, 1982, 1981 respectively in Figure 21a, 21b, 21c. In each figure, we include two education cohorts on both sides of the fake birth month cut-off. We collapse the residuals at birth quarter level after regressing *Identity* on survey year fixed effects. Each marker represent the simple mean of the residuals in each birth quarter. Fitted lines are derived from regression of dots using linear polynomials of birth quarter interacted with the dummy variable *TextbookExposure*.

## E Hoklo Ethnicity Distribution in Taiwan

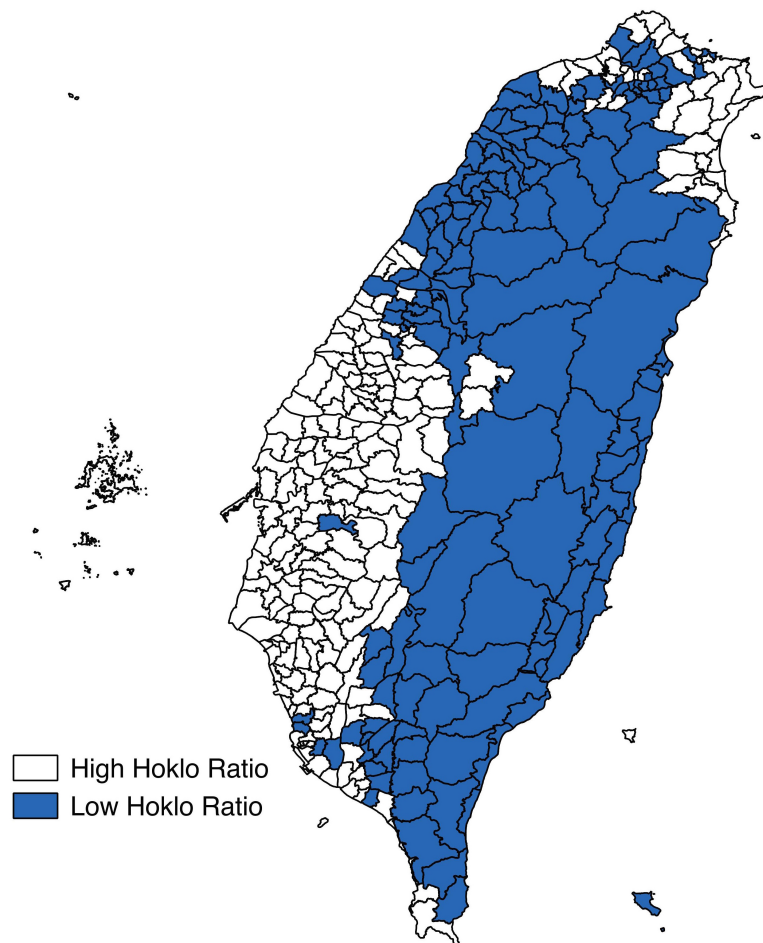


Figure 22: The Map with High/Low Hoklo People Proportion at Town Level

*Notes:* We compute the proportion of Hoklo people in each town and the population median of the proportion of Hoklo people (with 2004 population as weight) using the data from National Hakka Population Basic Information Survey Research surveyed in 2004. The median is 77.1%. *High Hoklo Ratio Towns* has proportion of Hoklo people higher than 77.1%; otherwise, the town belongs to *Low Hoklo Ratio Towns*.