

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

Wei-Lin Chen*, Ming-Jen Lin†, Tzu-Ting Yang‡

November 4, 2018

Abstract

This paper examines the causal effect of textbook content on individuals' national identity, exploiting a curriculum reform that introduced a new perspective on Taiwan's history for students entering junior high school after September 1997. Using a repeated nationally representative survey and a regression discontinuity design, we show that students exposed to the new textbooks were more likely to consider themselves Taiwanese. The effect is greater for academic track students and those living in neighborhoods where less people hold the Taiwanese identity. Finally, our results suggest the effect of textbook content on individuals' identity is *not* persistent in the long run.

*Corresponding author, Department of Economics, University of California San Diego, Email: wec155@ucsd.edu

†Department of Economics, National Taiwan University

‡Institute of Economics, Academia Sinica, Email: ttyang@econ.sinica.edu.tw.

1 Introduction

The more homogeneous the people, the easier it is to manage a nation. As a result, state leaders are incentivized to use the education system as an instrument for cultivating national identity—an essential step toward nation building. Empirical evidence has shown that, in the past 150 years, investments in mass education by governments have appeared in response to military threats, when patriotic people are required to prepare for future wars (Aghion et al., 2018). The causal effect underlying the intuition, and the transmission mechanism behind the effect of education on national identity formation, however, lack detailed scrutiny. National identity trends in society and cohort effects pose challenges to the identification of educational content effects. Specifically, these two effects interact with each other in the sense that students in different cohorts experience different levels of societal development. Furthermore, the introduction of new curricula may coincide with other social and political events influencing people’s national identity.

In this paper, we overcome these difficulties by exploiting a junior high school curriculum reform in Taiwan. In September 1997, the Taiwanese government published its *Knowing Taiwan* series of textbooks for social subjects: history, geography, and society. The history curriculum, in particular, adopted a new perspective on the nation’s history, and provided abundant Taiwan-related content, all of which had been absent from previous textbooks. The education system in Taiwan mandates that children born after 1 September must enter the education system the following year, such that people born in August will enter earlier than those born in September. This means that those born in September 1984 would have been the first month’s cohort to have studied the *Knowing Taiwan* series (i.e. They are 13 years old in 1997), while those born in August 1984 would have studied the old textbooks. These institutional features give us a unique opportunity to identify the causal effects of the junior high school curriculum (i.e. textbook content) on people’s national identity in later life, since those born either side of the cut-off would have experienced similar social events and political developments. In addition, the birth timing decisions of parents are unlikely to have been affected by this reform. Therefore, we can isolate any curriculum effect from other confounding factors by comparing the national identities of those born just before and just after September 1984, using a regression

discontinuity design. We measure national identity using a self-reported identity question from a repeated nationally representative survey—the Taiwan Social Change Survey. This survey has consistently asked respondents about their national identity through the question “Do you consider yourself Taiwanese, Chinese, or both?” over a long period of time.

We obtain three key findings from our research. First, our results suggest that students who studied the *Knowing Taiwan* series are on average 18 percentage points more likely to report themselves as Taiwanese than either Chinese or both (Taiwanese and Chinese) when they are around the age of 20 (18-23 years old), with a control group mean of 61%. Based on our estimates, we can calculate the persuasion rate using the formula employed in media economics ([DellaVigna and Gentzkow, 2010](#)). The estimated persuasion rate is 45%, thereby suggesting a large curriculum impact on people’s identity.

Second, we investigate the possible mechanisms through which school curricula can affect an individual’s national identity. Our subgroup analysis suggests the curriculum effects only appear in academic track students, who generally put more effort into studying textbook materials. This result implies that memorizing textbook content is a possible channel. In addition, we find that the curriculum effect varies substantially by an individual’s prior familiarity with Taiwan-oriented ideology, for which we use hometown ethnic distribution as a proxy. The four major ethnicities living in Taiwan are Hoklo, Mainlanders, Hakka, and Aborigines. In older generations, the Hoklo people are typically much more likely to have a Taiwanese identity than other ethnicities. We divide people according to their living environment, using the proportion of Hoklo people at the township level, and dividing the group into those with high and low proportions. We find a large curriculum effect in people living in towns with a smaller Hoklo proportion, whereas no effect is found among those living in towns with more Hoklo people. The result aligns with the predictions made by “belief-based (Bayesian learning) models,” in that people with weaker prior belief are more likely to be persuaded by new information ([DellaVigna and Gentzkow, 2010](#)).

Finally, we study the long-term effects of junior high school curricula on people’s national identity around the age of 30 (24-33 years old). Our results suggest that textbook contents have little impact on individuals’ identity in the long run. People who studied the old textbooks hold

a similar level of Taiwanese identity as those who studied the new textbooks. Thus, it appears that the Taiwanese identities of people who were not exposed to the new textbooks have *caught up* with those who were exposed to it.

Recently, a few papers have examined the effect of education policy on political institution preferences or ethnic identity. For instance, [Clots-Figueras and Masella \(2013\)](#) and [Fouka \(2015\)](#) studied the impact of language usage promotion and prohibition in the education system, respectively. [Clots-Figueras and Masella \(2013\)](#) found that changing from single-language (Spanish) to bilingual (Spanish and Catalan) education in Catalonia provided students with a stronger sense of Catalan belonging, which led further to changes in political party preferences in elections. [Fouka \(2015\)](#), on the other hand, found a backlash effect of German language prohibition in the US education system during 1917 and 1923. The author documented that children of German immigrants who experienced the language prohibition were more likely to marry Germans, choose more ‘German’ first names for their children, and be less likely to volunteer in World War II. She ascribed this backlash to the substitute property between parents’ investment in their children’s identity and schools’ influence in this regard. [Bai \(2018\)](#) examined the long-term effect of education under the authoritarian regime in Taiwan. He found that one additional year of exposure to authoritarian education during youth could substantially affect an individual’s political behaviors, such as preference for democracy or voting for an authoritative party.¹

Our paper is most relevant to [Cantoni et al. \(2017\)](#), who conducted a survey of students at Peking university (elite students) to study the effect of new curricula in China, adopted over different years in different provinces, ranging from 2004 to 2010. They found that the new curricula effectively shaped students’ ideologies according to the goals of the curricula; specifically, students exposed to the new curricula showed more trust in government and more skepticism toward unconstrained democracy and unconstrained free markets. The authors did not, however, find curriculum effects on ethnic identity. The main goal of the curricula in this study was to make students feel more integrated and have a united identity.

¹Our paper is also relevant to the political preference formation and culture transmission literatures. Previous literature has shown that political regimes and the media affect people’s policy and party preferences ([Alesina and Fuchs-Schundeln, 2007](#); [DellaVigna and Kaplan, 2007](#)).

Our paper stands apart from previous literature in the following ways. First, we provide the first evidence in support of the notion that, other than language usage, school curricula (i.e. textbook content) as an educational instrument—or more broadly a kind of authoritative media—does indeed influence national identity. Second, and different from [Cantoni et al. \(2017\)](#) who focused on elite students, we examine the impact of junior high school curricula on the more general population, using nationally representative survey data. Our estimates show that the curriculum effect exhibits heterogeneity, as far as effort put into studying textbook content and prior familiarity is concerned. Third, we contribute to the literature by investigating the long-term effects of curricula. Our results suggest the curriculum effect might *not* be permanent, and we thus have to be cautious about interpretation when observing the effect of curriculum policy in the short term.

The paper proceeds as follows. In section [2](#), we discuss the background of the curriculum reform and analyze the differences between the old and new curricula. Section [3](#) describes the data and sample used in this paper. Section [4](#) discusses our identification strategy—regression discontinuity design. Section [5](#) presents our main results. We then explore potential mechanisms through a subgroup analysis in section [6](#) and long-term effects in section [7](#). Finally, section [8](#) concludes.

2 Policy Background

2.1 The Curriculum Reform of the *Knowing Taiwan* Series

In 1994, the Taiwanese government announced a new curriculum for junior high school social subjects: history, geography, and society. The major change lay in the design of the first-year content. In earlier textbooks, Taiwan-related content accounted for only a small proportion of the text and was scattered through different volumes. However, the new curricula, especially history subject, aimed to provide not only 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 government published the new textbooks, known as the *Knowing Taiwan* series, and students entering junior high school in September 1997 were expected to utilize them

accordingly.²

The reform was comprehensive. Students across Taiwan who entered junior high school after September 1997 would study the *Knowing Taiwan* series. Though the major changes applied mainly to first-year textbooks, the second and third-years textbooks were also adjusted. Senior high school/vocational school examinations for students born after September 1984, compared to the examinations for earlier cohorts, were therefore based on different textbooks for all three years, thus ensuring that earlier education cohorts were not exposed to the *Knowing Taiwan* series. Here, we define the education cohort as students entering the compulsory education system in the same year, and we label them with the year they entered junior high school. For example, the 1997 education cohort entered junior high school in September 1997. They were the first to study the *Knowing Taiwan* series and were born between September 1984 and August 1985.

2.2 Comparison between Old and New Curricula

This curriculum reform aroused politicians' attention, because it brought to awareness the stark differences between two imagined nationalities, namely Chinese consciousness and Taiwanese consciousness (Liu et al., 2005; Wang, 2001). In particular, the new history textbooks moved away from the "China-oriented" angle seen in earlier textbooks to a "Taiwan-oriented" approach. In general, there are two main differences between the old and new textbooks: (1) the amount of content about Taiwan and (2) the context given about the relationship between Taiwan and China. Therefore, the new history textbook may have cultivated Taiwanese identity in two ways: First, there may have been a priming effect, due to students reading the word "Taiwan" more often. Second, the distinction made by describing Taiwanese and Chinese history separately, may have provided students with different information to associate with the two imagined groups, and hence helped them differentiate between Taiwanese and Chinese.

²This reform aroused fierce debate among political parties on whether the books were "appropriate". Political parties at that time were divided into two groups, with the likes of the Kuomintang and the New Party following the "successor to China" ideology, while the Democratic Progressive Party advocated "Taiwan independence" and considered the Kuomintang government, which had ruled Taiwan since 1945, a foreign regime. Discussions at the time, about whether the history textbook of *Knowing Taiwan* series 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 1997, 341 articles (five articles every day on average) about *Knowing Taiwan* appeared in the nation's four main newspapers.

2.2.1 A Substantial Increase in Taiwan-Related Content

Under the old curriculum, junior high school students studied the history of China for a year and a half, and then the history of the world for another year and a half, whereas under the new curriculum they studied the history of Taiwan in the first year (i.e. the history textbook of the *Knowing Taiwan* series), the history of China in the second year, and world history in the third year. In other words, the content on the history of China and the rest of the world in the old version was condensed in the new version so that new materials about Taiwan could be added.

In terms of time, teachers utilizing the new textbooks might have spent much more time on the history of Taiwan than previously. Under the old curriculum, teachers spent three semesters on the history of China (25 chapters) and only one chapter and a section was related to Taiwan.³ Assuming that teachers spent the same amount of time on each chapter and section in a volume, we approximate that they would have spent less than one-fifth of a semester on the history related to Taiwan. In contrast, the *Knowing Taiwan* history volume was designed to cover two semesters, with 116 pages of content. For comparison, the old textbooks contained only 16 pages dealing with Taiwan.

The notable addition of Taiwanese history was emphasized by the textbook's editors:

This book aims to introduce students to *the history about how ancestors of different ethnic groups made developments in Taiwan*. As a result, students would be expected to cultivate a cooperative spirit, patriotic feelings, and worldwide horizons. Also, it is hoped this will augment their understanding of *Taiwanese cultural assets*, and make them appreciate and treasure them accordingly.⁴

The intention of acquainting students with Taiwanese development was not apparent in the old version. This can be seen from the editors' preface to the old textbook on the history of China:

The history of China describes *the evolution of Chinese nationality, the change of*

³In the old textbook series, these 25 chapters were spread across three volumes, i.e. one volume per semester. Only a section in the 15th chapter, entitled "The rebellion of Koxinga against the Qing Dynasty and the development of Taiwan," and the 25th chapter, entitled "The achievement and vision of a base for revival," included Taiwan-related content.

⁴Emphasis in this paragraph is added by the authors.

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 patriotic feelings and a cooperative spirit, and to understand the nation's traditions, its position and the responsibility of the population.

In the new textbooks on the history of China, approximately the same words appear in the editors' preface, except that the part referring to "the evolution of Chinese nationality, the change of the territory" has been deleted.

2.2.2 Distinguishing between Taiwan and China

The new textbook not only contained a substantial increase in content about Taiwan, but also clearly distinguished between the concepts of Taiwan and China, in a contextual change. In general, the new textbook treated the history of Taiwan as a distinct entity from the history of China. In contrast, the old textbook did not emphasize this difference. Furthermore, depending on the context, the old textbook sometimes used "our country" to refer to China but sometimes also to refer to Taiwan. Thus, studying the old textbook could have made students confused about their national identity.

In their first grade of junior high school, students studying the old textbooks started learning the history of "our country (i.e. China)" through the statement that the earliest human beings lived in "our country (i.e. China)," namely *Homo erectus pekinensis*, in the Palaeolithic age. The "common ancestor" of *Chinese nationality* was Huang Di and the first dynasty of "our country (i.e. China)" was the Xia dynasty. The history of "our country (i.e. China)" therefore proceeded through sequential dynasties, from Xia to Qin, to Tang, and all the way to Qing.⁵ Interestingly, the old textbook also used "our country" to refer to Taiwan when it mentioned the development of the Kuomintang government in Taiwan after the 1949 Chinese Civil War (i.e. the Kuomintang-Communist Civil War).

In contrast, the term "our country" is used less in the history textbook of *Knowing Taiwan*

⁵Between the Ming and Qing dynasties in this straightforward development line, students saw the first appearance of "Taiwan," identified by the editors as a basis for Koxinga's fight against the Qing regime. It is worth noting that Koxinga is written as "recovering" Taiwan from the Dutch. The usage of the verb demonstrates explicitly the ideology behind the old textbook, showing that the editors viewed the ruling Dutch in the 17th century as a "foreign regime." Simultaneously, this implicitly claimed Taiwan as the territory of "our country (i.e. China)" before Dutch rule.

series or for the textbook on the history of China in the new curriculum. “Taiwan” and “China” are used instead. More precisely, “our country” only appears in descriptions of Taiwan. Following the divided usage of terms, Taiwanese history stands out not as part of the history of China but as an individual entity in the new history textbooks.

To sum up, in the old curriculum, “Taiwan” was virtually ignored and “our country” usually referred to “China”. In the new curriculum, “Taiwan” and “China” have been explicitly separated so that readers have the chance to distinguish between them.

3 Data and Sample

3.1 Data

The data used in this paper were taken from the Taiwan Social Change Survey (TSCS), which is a nationally representative repeated cross-sectional survey for respondents aged above 18 in Taiwan. The sample size of each TSCS wave is around 1,800 to 2,200 respondents.

Three features of the TSCS make it suitable for our analysis. First, it asks respondents consistently about their national identity through the following question:

- In our society, some people call themselves Taiwanese, some Chinese, and some both.
Do you consider yourself Taiwanese, Chinese, or both?

This feature allows us to combine different survey waves, in order to compare the short- and long-term impacts of curriculum reform on national identity. Secondly, the TSCS records the birth year *and* birth month of respondents. Since the school year in Taiwan starts in September, by exploiting this feature, we can identify the correct educational cohort, which is crucial for our regression discontinuity design. Third, the TSCS holds rich demographic information about respondents, which helps us to further investigate the mechanism through subgroup analysis.

3.2 Sample

The first educational cohort exposed to the *Knowing Taiwan* reform, born in September 1984 or later, was first surveyed in 2003. To balance out the regression analysis respondents before and after the reform, we hence include surveys held from 2003 onward, which contain the national

identity question and enough background information: These are the 2003, 2004, 2005, 2009, 2010, 2012, 2013, 2014 and 2015 waves.⁶

We drop any respondents who reported being born outside Taiwan and those who reported that the place they had lived the longest before they were 15 was outside of Taiwan, since we could not be sure that they had entered junior high school and hence been exposed to the curriculum reform. In addition, we drop respondents whose answer to the national identity question was “Other.” These selection rules remove 2% of the main regression sample (i.e. short-term sample). The main results in this paper are not influenced by the sample selection.

3.3 Outcome Variable and Covariates

Based on the TSCS’s national identity question, we create the outcome variable as a dummy variable *Identity* by assigning one to respondents answering “Taiwanese” and zero to those answering “Chinese” or “Both.”⁷ In our main regression sample, only 3.8% of respondents answer “Chinese,” indicating that, in this generation, very few people identify as exclusively Chinese. Most of the respondents have an exclusively Taiwanese identity (64.5%) or a dual identity (31.7%), considering themselves to be both Chinese and Taiwanese.

We include the respondent’s gender, their parents’ education level, and their parents’ ethnicity in the analysis. The rich demographic variables help us construct different subgroups to explore the possible mechanisms of the curriculum effect. In particular, we focus on the respondents’ education track and the ethnic group distribution in the respondents’ hometown. Regarding the respondents’ education track, we split the sample into an academic track and a vocational track. The respondents whose final educational attainment is senior high school or university are assigned to the academic track. Those whose attainment is junior high school,

⁶Note that the TSCS held two waves in 2014. In 2009, and in one of the waves in 2014, the survey did not request the demographic information we need in some of our regressions. Hence, we do not include these two waves in the regressions that control for demographic variables.

⁷Since the measurement of national identity is based on a self-reported response, the natural question is: Does this measurement truly reflect respondents’ national identity? One possible explanation for a change in *Identity* (if observed) is that previous students were afraid to respond that they felt Taiwanese. The new textbooks provided not only a Taiwanese identity, but also the message that viewing oneself as Taiwanese was no longer taboo. We provide two counterarguments to this explanation. First, the simple mean of *Identity* for the control group in our main analysis sample is 0.6. When over half of one’s peer group identify themselves as Taiwanese, it is hard to believe that the Taiwanese identity is indeed taboo. Second, the change in *Identity* should be visible in different subgroups if this explanation is true, but in section 6 we find this is not the case.

vocational high school, or vocational university are assigned to the vocational track.

For the ethnic distribution in the respondents' hometown, we use a question from the TSCS that reads as follows: "Where did you live longest before you were 15 years old?" The responses are on the township (zip code) level, and so we regard them as reflecting where the respondents lived when in junior high school (i.e. their hometown). This hometown information is combined with township-level ethnicity data to approximate how many Hoklo people the respondents were surrounded by in their daily lives before junior high school. Compared to other ethnic groups, Hoklo people are more likely to consider themselves Taiwanese because they migrated to Taiwan from Southern China a long time ago (i.e. around 400 years ago).

The ethnicity data come from National Hakka Population Basic Information Survey Research, conducted in 2004 with a sample size of 37,693, equivalent to about 100 people in each town. We use the responses to the question: "You consider yourself as..?" The six options included (1) Taiwan Hakka, (2) Mainland Hakka, (3) Hoklo, (4) Mainlander, (5) Aborigine, and (6) Foreigner. 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 (*Hoklo Proportion*) in the town in question. We then estimate the population median of *Hoklo Proportion* in Taiwan, using the value for each town, weighted by the total population of the town in 2004 (population data sourced from the Ministry of Interior). The median is 77.1%.⁸ Finally, we construct the dummy variable *Less-Hoklo-Hometown* by assigning one to respondents whose hometown's *Hoklo Proportion* is less than 77.1%, and zero to all other respondents. If the respondent lived in a town with a *Hoklo Proportion* below the median, he was less likely than the average person in Taiwan to have met a Hoklo person in his daily life.

4 Empirical Specification

4.1 Graphical Evidence

Figure 1 plots the simple mean of *Identity* in each educational cohort, using all available data. We observe a roughly 10% increase in Taiwanese identity between the 1996 and 1997 education cohorts (i.e. between the last to study the old curriculum and the first to study the *Knowing*

⁸The mean using the same data and weights is 73.2%

Taiwan series). Two important caveats should be noted in the above analysis. First, compared to people who enter school earlier, those who enter school later are less likely to have been surveyed in early years since they are too young to become respondents. In addition, people's national identity might be affected by social events happening in the survey year, so the above change in Taiwanese identity could be confounded by survey year effects. Second, the result in Figure 1 might be mixing up the short-term and long-term effects of the school curricula on Taiwanese identity. Since we use all available survey waves from 2003 to 2015 to plot Figure 1, this implies that some in the sample would have been surveyed in the early stages of their life and some in the later stages. To alleviate the above concerns, we control for the survey year fixed effect and restrict our sample to fewer education cohorts in the rest of our analysis, namely those born between September 1982 and August 1986 (four education cohorts, two of which would have studied the *Knowing Taiwan* series). In addition, we first analyze these cohorts when they were relatively young, aged from 18 to 23 and surveyed from 2003 to 2005 (henceforth **short-term sample**). To examine if the curriculum effect is persistent, we examine the same education cohorts surveyed from 2009 to 2015, when aged between 24 and 33 (henceforth **long-term sample**).

4.2 Regression Discontinuity Design

Different cohorts of students would have been exposed to different societal trends, which may have affected their national identity formation. Thus, we use a regression discontinuity design (RDD) to eliminate this problem by comparing the identities of people born close together (i.e. around September 1984). The reason this works is that close birth cohorts should experience almost the same societal developments while growing up. The major difference is that those born just after September 1984 would have studied the *Knowing Taiwan* series, while those born just before September 1984 would have studied the old textbooks. At first glance, we should conduct an RDD on an education cohort (i.e. academic year) basis, since the treatment status varies at that level. However, people in the same education cohort may have experienced different events that could have altered their national identity. An example of this relates to voting.

Elections in Taiwan are generally held in December, January, and March, and the age at which one becomes eligible to vote is 20. In some elections, people born in the first half of the education cohort would have been eligible, while those born later would not have been. Students in the first cohort exposed to the curriculum reform offer one example in this regard. The sixth legislative election was held on December 11, 2004, splitting the education cohort into two groups: people who had the voting right (born before December 11, 1984) and people who did not have it (born after December 11, 1984). Students in the last cohort studying the old textbook provide another example. The event in this case was the presidential election that took place on March 20, 2004. The reason this is important is that politicians in Taiwan debate fiercely on the subject of national identity in elections. Thus, different “first vote” experiences may affect people’s national identity formation. Bearing in mind such differences embedded in respondents within an education cohort, we measure birth cohort at month level and estimate the following regression::

$$Identity_{ijt} = \alpha_0 + \alpha_1 TextBook_i + f(m; \beta) + \gamma X_i + \eta_j + \delta_t + \epsilon_{ijt} \quad (1)$$

where $Identity_{ijt}$ indicates the dummy variable defined in section 3, for individual i , with hometown j , interviewed at time t . The variable $TextBook$ indicates whether the respondent was exposed to the curriculum reform and takes the value one if the respondent reported himself born after September 1984, and zero otherwise. We use birth cohort measured by month as our running variable, and center it at September 1984, the first month affected by the reform. In our main specification, we estimate equation (1) within a bandwidth of 24 months before and 24 months after September 1984 (i.e. use the sample born between September 1982 and August 1986). In addition, we specify $f(m; \beta)$ as a linear function, but allow the slope to be different on either side of the cut-off. That is, $f(m; \beta)$ is the first-order polynomial of birth cohort m interacting fully with $TextBook$.⁹ In a later section, we examine whether our main results are sensitive to the bandwidth choices and different specifications.

Our primary interest is in α_1 , which measures any deviation away from the continuous relation between the birth cohort and Taiwanese identity $Identity_{ijt}$ at the cut-off (i.e. when

⁹We also include second-order polynomial of birth cohort m interacting fully with $TextBook$ for a robustness check.

the treatment variable *TextBook* switches from 0 to 1). If the only disparity around the cut-off is the year of entering junior high school, which consequently includes our treatment, the studying of the *Knowing Taiwan* series in junior high school, α_1 represents the causal effect of doing so on Taiwanese identity.

In order to single out the overall effect of societal trends in each survey year, we include the survey year fixed effect (δ_t) in all specifications. In some specifications, we also include the hometown fixed effect (η_j), which helps us control for regional factors possibly influencing national identity formation, such as local support for a certain political party. To increase the precision of our estimates, we also include demographic variables (X_i) which might influence national identity formation, including gender, parents' education, and parents' ethnicity.¹⁰ The parents' ethnicity and education level capture the family's influence on the respondents' national identity.¹¹ Since, in a later section, we find that the share of Hoklo people in the respondents' hometown (*Less-Hoklo-Hometown*) exhibits a discontinuity at the cut-off (see the last column of Table 3), we also include that variable in our regression to control for its effect. ϵ_{ijt} is an error term that reflects all of the other factors that affect the outcome variables. Finally, standard errors are clustered at the birth cohort level (i.e. birth year-month).

5 Results

5.1 The Effect of the *Knowing Taiwan* Series on Taiwanese Identity

Figure 2 displays relationship between Taiwanese identity and birth cohort. We group up the sample by every three birth year-months to increase sample size of each dot. Thus, each dot in Figure 2 represents average of variable *Identity* (i.e. Taiwanese identity) by three birth year-month cohorts (i.e. the birth year-quarter cohort), after it has been regressed on the survey year

¹⁰The definitions of the demographic variables (X_i) are as follows: (1) Gender: female being 1, male being 0. (2) Father/Mother ethnicity: Hoklo fathers/mothers being 0, otherwise 1. (3) Father/Mother education level: father/mother with no education, elementary school, and junior high school education level being 0, otherwise 1.

¹¹The empirical literature has shown that parents will invest in strengthening their offspring's national identity, to make it the same as their own (Fouka, 2015). In the older generation in Taiwan, ethnicity largely determines one's national identity; for example, Mainlanders are much more likely to identify as Chinese than other ethnic groups. We define Mainlanders as people at least one of whose parents is a Mainlander; the rest, we define as Others. Combining the 2003, 2004, and 2005 TSCS waves, and retaining only those respondents born between 1950 and 1964, we find that the proportions of Mainlanders reporting themselves as Taiwanese, Chinese, and both are 25.9%, 10.6%, and 61.4%, respectively, while the corresponding figures for the Others are 67.2%, 2.7%, and 28.7%

dummies (i.e. controlling for the survey year fixed effect).¹² The lines in Figure 2 represent fitted regressions of the cell’s mean dots, using first-order polynomials interacting with the dummy variable *TextBook*. By doing this, we can eliminate the potential confounding effect of the survey years. The fitted line in Figure 2 suggests that the discontinuity of *Identity* is roughly 20% around the cut-off.

Table 1 shows the regression results of estimating specification (1). The first-order polynomials of birth cohort *m* fully interact with *TextBook*, and the survey year fixed effects are included in all regressions. Column (1) reports our baseline results. Consistent with the graphical evidence in Figure 2, the estimate of the coefficient on *TextBook* is 0.16 and statistically significant. In other words, studying the new textbook (i.e. the *Knowing Taiwan* series) can increase one’s probability of reporting oneself as Taiwanese by around 16 percentage points. Note that the mean of *Identity* on the left side of the cut-off (the control group) is around 61%.¹³

In columns (2) to (4), we gradually include hometown fixed effects, demographic variables, such as gender, parents’ education level, and parents’ ethnicity, and *Less-Hoklo-Hometown* to increase the precision of the estimates and lessen the potential bias due to discontinuities in observables at the cut-off. In general, we find qualitatively similar estimated coefficients on *TextBook* across the different specifications. Our results suggest that studying *Knowing Taiwan* does significantly increase the likelihood of having a Taiwanese identity by around 18 percentage points.

5.2 Discussion: Persuasion Rate

In this section, we provide the persuasion rate, calculated by the formula used in media economics (DellaVigna and Gentzkow, 2010):

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

¹²The graph is at the birth year-quarter level, so first dot in Figure 2 represents average *Identity* (after controlling for the survey year fixed effect) for those born in September, October, and November 1982 and the last dot represents average *Identity* (after controlling for the survey year fixed effect) for those born in June, July, and August 1986. In the later sections, we use similar way to display Figure 4, Figure 6, and Figure 7.

¹³This is the mean of *Identity* across all those in the sample who were born between September 1982 and August 1984 (two education cohorts).

where e_i denotes the share of group i receiving the message (the textbook content in our case), and y_i the share of group i adopting the behavior (i.e. considering themselves Taiwanese in our case). The subscripts t and c represent the treatment and control groups. The persuasion rate measures the degree to which the treatment persuades people to adopt the behavior, scaled by the share of people receiving the messages and the share of the control group “to be persuaded” ($1 - y_c$). In our case, since all students born after September 1984 were exposed to the new textbook, $e_t - e_c = 1 - 0 = 1$. The persuasion rate reported in the first column in Table 1 is calculated as $100 \times \frac{0.176}{1} \times \frac{1}{1-0.608} = 44.8$. This 45% persuasion rate is quite high compared to the persuasion rates found in the media economics literature, which are barely higher than 20%. Our estimate, however, aligns with the persuasion rate found in [Cantoni et al. \(2017\)](#), in which paper more than a quarter of the persuasion rates were higher than 20%, and the highest was 41%. The high persuasion rate is not that surprising after taking into account the degree of exposure: Students have to have studied the *Knowing Taiwan* series for at least a year, and they also have to have spent three years memorizing the materials for the high school admission examinations. This exposure is much greater than typically occurs with specific TV or radio programs.

5.3 Robustness Check

We validate the robustness of the main results in two ways. First, we discuss their sensitivity to different empirical settings, such as the inclusion of higher polynomial orders and the choice of bandwidth. Second, we investigate the validity of the identification assumption for RDD, by examining the smoothness of observable covariates and conducting a series of falsification tests.

5.3.1 Choices of Polynomial Order and Bandwidth

To examine whether our results are sensitive to different parametric specifications, Table 2 displays estimates based on a specification with second-order polynomial (i.e. quadratic spline). The estimated results suggest studying *Knowing Taiwan* can on average increase Taiwanese identity by 18-25 percentage points, which is a range quite similar to our main estimates. Next, we examine the robustness of our estimates over a wide range of bandwidths. Figure 3 shows

the point estimates of the coefficient on *TextBook* and their corresponding 95% confidence intervals using the same specification as in column (4) of Table 1, with bandwidths ranging from two education cohorts (i.e. 24 months) to one (i.e. 12 months) on each side of the cut-off. The magnitudes of the point estimates remain similar as we narrow down the birth month window, showing that the results in Table 1 are not sensitive to the bandwidth choice.

5.3.2 Smoothness of Observable Covariates at Cutoff

A key identification assumption of RDD is that the individuals' characteristics should be similar on both sides of the cut-off (i.e. born in September 1984). In other words, no other confounding factors should change in September 1984. To investigate this issue, we examine whether the selected observable characteristics are balanced on both sides of the cut-off. We use these characteristics as outcome variables and estimate equation (1) without controlling for the covariates X_i and hometown fixed effects. The regression results are shown in Table 3. Most observable characteristics do not exhibit significant discontinuities at the cut-off—the only exception is the variable *Less-Hoklo-Hometown*. The last column of Table 3 suggests this variable exhibits a significant discontinuity, with a size of 0.155. In other words, it is significantly more likely that we will observe a respondent who lived in a town with fewer Hoklo people on the right-hand side of the cut-off. As mentioned before, in order to lessen any potential bias, we include this variable in the main specification and find that our estimates are robust to its inclusion.

5.3.3 Placebo Test

In this section, we further examine our identification assumption, that no other confounding factors change at the cut-off, by conducting a series of placebo tests. One potential confounding factor could be the mental age effect: People who were born on left-hand side (i.e. August) of the birth year-month cut-off would have been more mentally mature than those on the right-hand side (i.e. September), since they had entered the school system earlier and thus, at any given time, may have had more work or social experience, which might have affected their Taiwanese identity. That being the case, we should observe similar jumps in September for every birth cohort. To examine this hypothesis, we estimate equation (1) for three fake reforms.

We take 1996, 1995, and 1994 as fake curriculum reform academic years and thus treat

September 1983, 1982 and 1981 as fake birth year-month cut-offs.¹⁴ We then replicate the results in Table 1 for each fake curriculum reform, using the same TSCS waves in 2003, 2004, and 2005. Note that we only include two education cohorts (i.e. 24 months) on each side of the fake birth year-month cut-off, to make the falsification results comparable to our main results.

Table 4 shows the results of the falsification regressions. The estimated ‘treatment effects’ are generally insignificant and the magnitudes are quite small. Thus, the results of the above placebo tests suggest our main estimates might not be driven simply by the mental age effect or other confounding factors.

6 Mechanisms

In this section, we explore the possible mechanisms through which school curricula (i.e. textbook contents) might affect an individual’s national identity, by conducting subgroup analysis along two dimensions: education track and the ethnic distribution of one’s hometown. For each subgroup, we estimate equation (1) and conduct a similar RDD analysis to that seen in section 5.1.

6.1 Memorization: Subgroup Analysis by Education Track

One possible channel through which school curricula might affect one’s national identity is memorization. Students who paid more attention to studying the textbooks should be associated with higher treatment intensity, in the sense that they may have memorized more Taiwan-related texts. Specifically, we examine this mechanism by utilizing a subgroup analysis based on intensity of exposure to the new textbooks. The ideal proxy for this intensity is the grade achieved in the social subjects in the senior high school or vocational high school admission examination.¹⁵ Unfortunately, the TSCS data does not include such information, but we can distinguish roughly between high and low levels of effort devoted to academic subjects in general in junior high school by the students’ choice of education track.

After completing compulsory education, students in Taiwan are divided into two educational

¹⁴They are 13 years old in 1996, 1995, and 1994, respectively.

¹⁵These junior high school graduates in our sample, no matter which education track they wish to proceed with, take the same national examination and use the grade they achieve to apply for senior or vocational high school.

streams: the academic track and the vocational track. The choice of track is endogenous and highly correlated with the effort students put in to studying when in junior high school. Students who were motivated to pursue more academic knowledge would have studied the textbooks far more, to give them a better chance of being selected by their preferred senior high school. On the other hand, common wisdom suggests that parents in Taiwan encourage students who lack motivation but are adept at obtaining excellent grades (for example, memorize the material more quickly than the average person) to opt for the academic instead of the vocational track. Consequently, the education track implies something about the students' latent exposure to the *Knowing Taiwan* series.

We categorize the respondents into two groups. Those with senior high school or university as their final educational attainment level are labelled as belonging to the academic track, while the others, having junior high school, vocational high school, or vocational university as their final level of attainment, are labelled as belonging to the vocational track.

Figure 4 displays the relationship between Taiwanese identity and birth cohorts by the academic track (Figure 4a) and vocational track (Figure 4b) respondents. We observe a significant jump around the cut-off in Figure 4a but no systemic pattern in Figure 4b. We also note that, although no systematic variation is found in Figure 4b, the mean levels of the dots before and after the birth month cut-off are close together, hinting that the *Knowing Taiwan* series did not significantly affect Taiwanese identity formation for vocational students.

Table 5 presents RDD estimates based on equation (1) for academic track respondents (Panel A) and vocational track respondents (Panel B), respectively. The results from the academic track respondents suggest that studying *Knowing Taiwan* can significantly increase the probability of such students having a Taiwanese identity, by roughly 30 percentage points, a much larger effect than in the overall sample as shown in Table 1 (i.e. 18 percentage points). In contrast, the results from the vocational track are generally not significant.

One concern over splitting the sample by educational attainment is endogeneity, namely the possibility that introducing *Knowing Taiwan* altered the demographic composition of the students in the two tracks. Specifically, one might argue that students who originally had a stronger Taiwanese identity may have attained better academic grades than those with an ini-

tially weaker Taiwanese identity, and hence the former would have been more likely to enter the academic track. We posit, however, that this is implausible. According to the calculation of the overall score in the admission exams across different subjects, the *Knowing Taiwan* content accounted for only one-fifteenth.

This subgroup analysis complements existing evidence of a curriculum effect from [Cantoni et al. \(2017\)](#). Since [Cantoni et al. \(2017\)](#) conducted their survey at Peking University (i.e. an academic track school), their sample consisted of students who excelled at memorizing textbook materials. Thus, they could not tell whether the new curriculum had indeed influenced those who did not need to expend too much effort on studying the textbooks. Our results corroborate the notion that it did do so, suggesting that, in the case of textbook content, the degree of exposure matters in terms of the persuasion effect.

6.2 Prior Belief: Subgroup Analysis by Hometown Ethnicity Composition

According to “belief-based” models, people who possess less prior belief can be affected more by new information ([DellaVigna and Gentzkow, 2010](#)). We test this prediction by conducting subgroup analysis based on the following two groups: people who lived in towns with more Hoklo people than the median, and all others.¹⁶ This division distinguishes between students with more or less prior familiarity with Taiwan-related knowledge. Children may randomly pick up cultural ideas from parents or role models in the neighborhoods in which they live ([Bisin et al., 2011](#)). Since Hoklo people in the older generation typically have a stronger Taiwanese identity, the proportion of Hoklo people surrounding the respondents is likely to influence the probability that a junior high school student would have picked a role model with a strong Taiwanese identity. Furthermore, in terms of political discussions such as election campaigns, people living in towns with fewer Hoklo people would have been exposed to fewer Taiwan-oriented speeches, since politicians running for local elections have to cater to local people’s political preferences, including those related to national identity.

To show that our subgroup criterion distinguishes between local environments with different

¹⁶We use the variable *Less-Hoklo-Hometown*, which takes the value one if the proportion of Hoklo people in one’s hometown is less than the population median (77.1%), and zero otherwise, to divide the sample. We provide a map demonstrating the Hoklo ethnicity distribution based on the *Less-Hoklo-Hometown* in Appendix A.

national identity trends, we utilize the 1992, 1995, 1998 and 2000 TSCS waves and calculate the mean of *Identity* in high- and low-*Hoklo Proportion* towns in different survey years. Figure 5 suggests people living in high-*Hoklo Proportion* towns, on average, would be more likely to report themselves as Taiwanese than those living in low-*Hoklo Proportion* towns (i.e. around 10-15% more). This assures us that students living in these two kinds of area would have faced significantly different social environments in terms of issues regarding national identity when in junior high school and elementary school—the time when they would have absorbed this information from the environment in which they were living. We argue that the curriculum effect would have been greater for students living in low-*Hoklo Proportion* towns, according to “belief-based” models, since they would have been less familiar with Taiwan-related knowledge beforehand.

Figure 6 displays the relationship between Taiwanese identity and birth cohorts depending on respondents living in low-*Hoklo Proportion* towns (Figure 6a) and high-*Hoklo Proportion* towns (Figure 6b). For the former group of respondents, Figure 6a suggests there is a substantial increase in Taiwanese identity at the cut-off. However, for the latter group, we find little evidence of a change in Taiwanese identity at the cut-off (see Figure 6b). Consistent with the graphical evidence, Panel A of Table 6 suggests studying *Knowing Taiwan* can significantly increase the Taiwanese identity of respondents living in low-*Hoklo Proportion* towns, by 35 percentage points. In contrast, for respondents living in high-*Hoklo Proportion* towns, studying *Knowing Taiwan* has little impact on their Taiwanese identity (see Panel B of Table 6).

6.3 Relationship between Main Results and Subgroup Analysis

The above subgroup analysis displays significant heterogeneity in the curriculum effect along two dimensions: education track and proportion of Hoklo people in one’s hometown. The heterogeneity in curriculum effect helps further, in ruling out two concerns. The first is the measurement concern. One might suspect that the main results come from the idea that students in earlier cohorts (i.e. born before September 1984) were afraid of proclaiming their true national identity. If this were the case, then the subgroup analysis would suggest that only students in an academic track and those living in low-*Hoklo Proportion* towns were afraid to do so, which

is very unlikely. Second, for the same reason, if the main results originated from other factors, such as significant social events, those events should have affected students in all subgroups. We are thus, based on this subgroup analysis, more confident that the effect we detect in the main results derives from the introduction of the new *Knowing Taiwan* series.

7 Long-Term Results

Up to this point, we have established that the introduction of new textbooks, in the form of the *Knowing Taiwan* series, influenced students' national identity as stated at the age of 18 to 23 years old (short-term sample). The natural question to ask is whether or not the impact was transitory or persistent. Clots-Figueras and Masella (2013) measured Catalan identity and observed the voting behavior of their sample cohort when roughly 25 to 30 years old, while Fouka (2015) used ethnic intermarriage and first names of offspring to measure strength of ethnic identity. Both papers found significant language policy effects in such long-term outcomes. We explore this issue by examining the long-term sample (respondents surveyed from 2010 to 2015, when 24 to 33 years old).

Figure 7 displays the relationship between Taiwanese identity and birth cohorts for the long-term sample.¹⁷ We find the mean level of Taiwanese identity to be quite similar on either side of the cut-off. Consistent with the graphical evidence, the regression results in Table 7 suggest the coefficients of *Knowing Taiwan* are all small and insignificant across different specifications, which is quite a different finding from our main results. That is, in the long run, no matter which textbooks people read, they are similarly likely to hold a Taiwanese identity.

Based on our research design, there are two possible explanations for this finding. First, the likelihood of a Taiwanese identity among people who read the new textbooks (i.e. *Knowing Taiwan*) “retreats” to the original level (i.e. that in the control group) in the long run. Second, the likelihood of a Taiwanese identity among people who read the old textbooks “catches up” with that for those who read the new textbooks in the long run. Figure 8 displays the trend in Taiwanese identity for the treatment and control groups. We find that the level in the control group (i.e. the circle symbols) is similar to that in the treatment group (i.e. the diamond symbols)

¹⁷As in our main results, we measure birth cohorts at the year-quarter level and plot average *Identity* after controlling for survey year fixed effects.

during 2009 to 2015, when the respondents are around the age of 30.

One possible explanation for the convergence of Taiwanese identity across both cohort in the long run is peer interaction. Although students who studied the old textbook were originally endowed with weaker Taiwanese identity than students exposed to the new textbook (in the short run), the “old textbook” effect may have been lessened as a result of peer interaction, because the former students would have encountered more and more “Taiwanese-oriented” peers in their living environment. Another hypothesis is that the control group was affected by the overall increasing trend in Taiwanese identity. Given our research design and data, we cannot confirm which mechanism drives our long-term results.

8 Conclusion

This paper examines the causal effect of school curricula on national identity formation, exploiting the textbook reform which introduced new and far more Taiwan-oriented materials (the *Knowing Taiwan* series). We use a regression discontinuity design to tease out a national identity societal trend, and find people who studied the *Knowing Taiwan* series have a much stronger Taiwanese identity than those who studied the old textbooks. Our results suggest that educational content can influence not only preferences over political institutions ([Cantoni et al., 2017](#)), but also national identity, thereby providing more empirical evidence of the homogenizing effect of education on people, as advocated in the literature ([Alesina and Reich, 2015](#)).

We contribute to the literature by presenting three facts. First, the degree of exposure to educational content matters. Students who studied the new textbook to a greater degree, measured by the division into academic and vocational track in the Taiwanese education system, were affected more strongly by the new textbooks. Second, the educational content effect is greater for students with weaker prior beliefs (identified by hometown ethnic distribution), which is aligned with “belief-based” models in the persuasion literature. Whether people who hold stronger or weaker prior beliefs are more affected by education policies is not clear in the literature. [Voigtländer and Voth \(2015\)](#), for instance, found that people who held a stronger prior anti-Semitic attitude (identified by their support for anti-Semitic parties in the previous election) were affected more by anti-Semitic indoctrination between 1933 and 1945 (i.e. they

exhibited the largest increases in anti-Jewish attitudes). Why persuasion is effective in different subgroups under different contexts is a potential research question for the future. The above two findings, regarding the importance of degree of exposure and prior beliefs, generalize the findings in [Cantoni et al. \(2017\)](#), which documented the curriculum effect in Peking university students, a group of students potentially more affected by textbooks, due to their hard work in senior high school. Finally, we document that the curriculum effect is *not* persistent in the long run, by showing that textbook content has little impact on a person's Taiwanese identity at the age of 30.

This paper is subject to a few limitations. The biggest one is that we cannot investigate why textbook contents have little impact on people's national identity in the long run. Although we provide two potential explanations for this finding, exploring whether either of these is more aligned with the data will have to be left as a potential research direction for the future.

References

- Aghion, P., Jaravel, X., Persson, T., and Rouzet, D. (2018). Education and military rivalry. *Journal of the European Economic Association*, page jvy022.
- 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*.
- Bai, Y. (2018). Good bye chiang kai-shek? the long-lasting effects of education under the authoritarian regime in taiwan. *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. (2017). Curriculum and ideology. *Journal of Political Economy*, 125(2):338–392.
- 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.
- 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.

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.

Figures

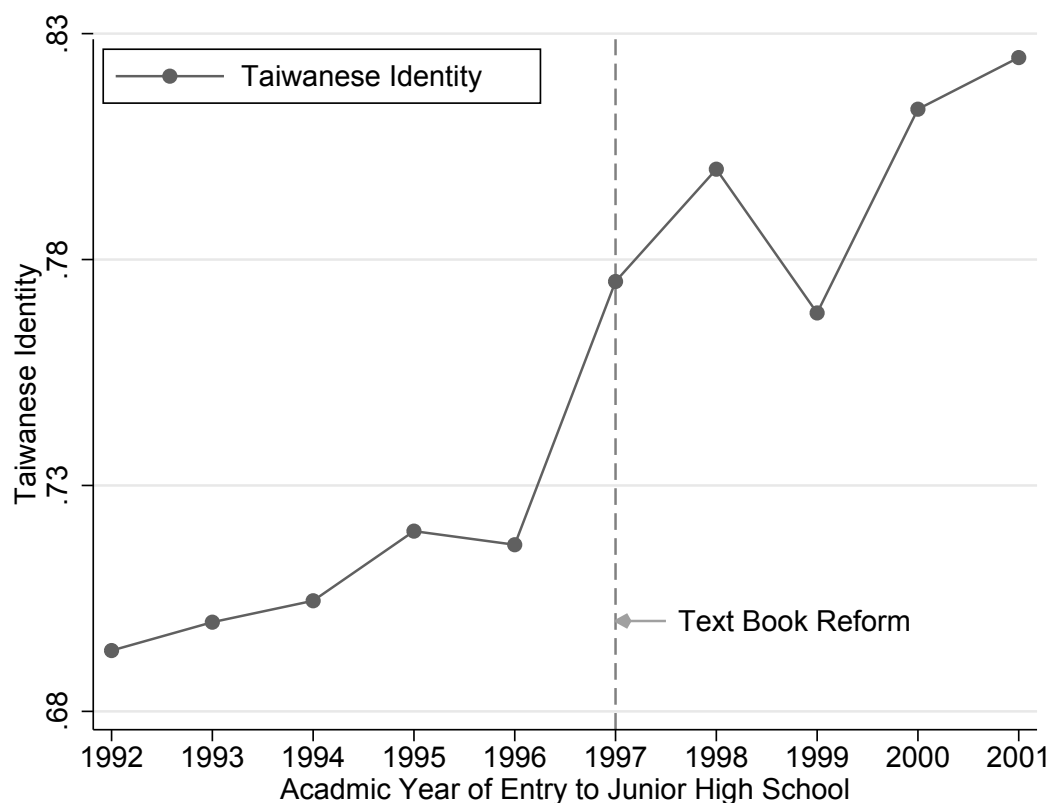


Figure 1: Taiwanese Identity and Education Cohorts

Notes: We pool all available TSCS data (i.e. 2003, 2004, 2005, 2009, 2010, 2012, 2013, and 2014 waves) and include education cohorts from 1992 to 2001. We include the 2009 wave and one of the two waves in 2014, which are not included in some of our regression analysis, since we do not require demographic information to draw the graph. Taiwanese identity is measured by a dummy variable *Identity*. It assigns one to respondents answering “Taiwanese” and zero to those answering “Chinese” and “Both”. Each dot represents average Taiwanese identity (*Identity*) for specific education cohorts.

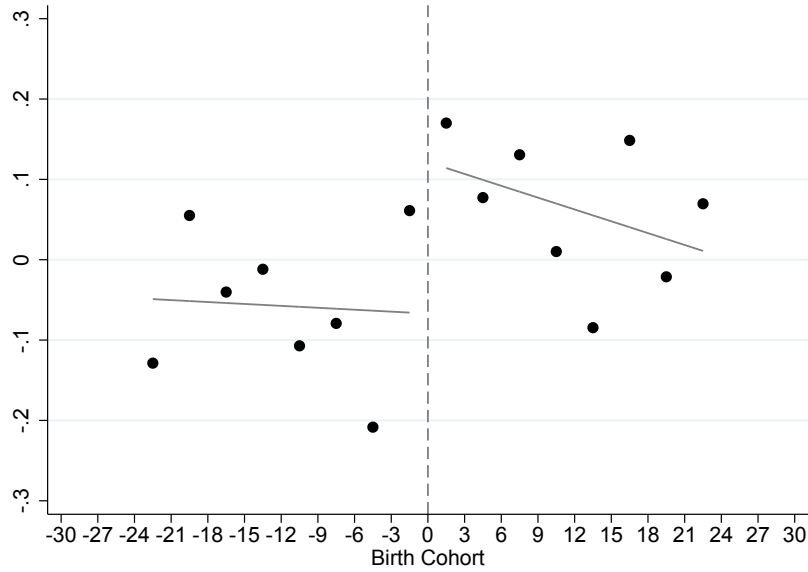


Figure 2: Taiwanese Identity and Birth Cohorts: Main Results

Notes: We pool data from 2003, 2004, 2005 TSCS and use the sample born between September 1982 and August 1986. We first regress *Identity* on survey year dummies and then collapse the residuals at birth year-quarter level (i.e. three birth year-month cohorts) to derive the dots. Thus, the first dot in this figure represents average *Identity* (after controlling for the survey year fixed effect) for those born in September, October, and November 1982 and the last dot represents average *Identity* (after controlling for the survey year fixed effect) for those born in June, July, and August 1986. Fitted lines are from regression of the dots on a first order polynomial of birth year-quarter interacted with *TextBook* dummy variable.

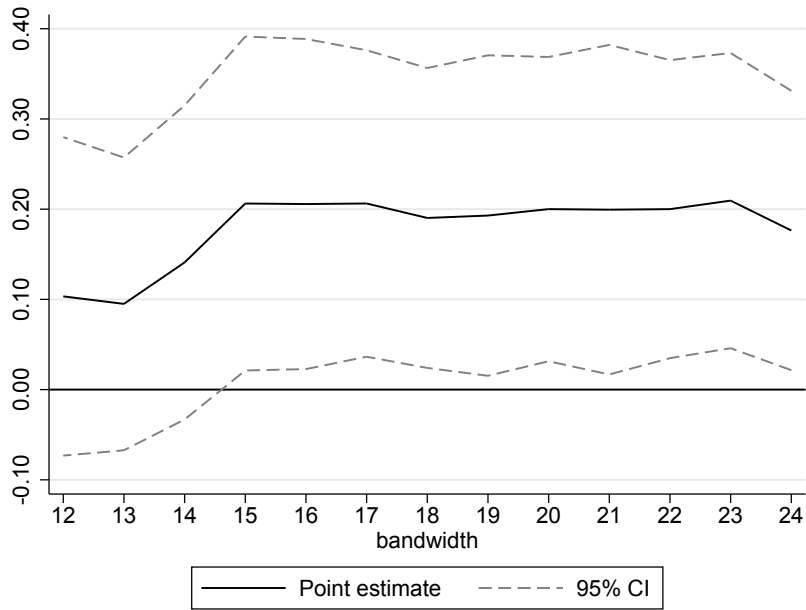


Figure 3: RD Estimates across Different Bandwidth Choices

Notes: We run regressions as column (4) in Table 1 with different bandwidths: 12 to 24 months on each side of the cut-off, i.e., two education cohorts. The solid line represent the point estimates of coefficients on the *TextBook* dummy variable and the dotted line represents the corresponding 95% confidence interval derived from standard errors clustered at birth year-month level.

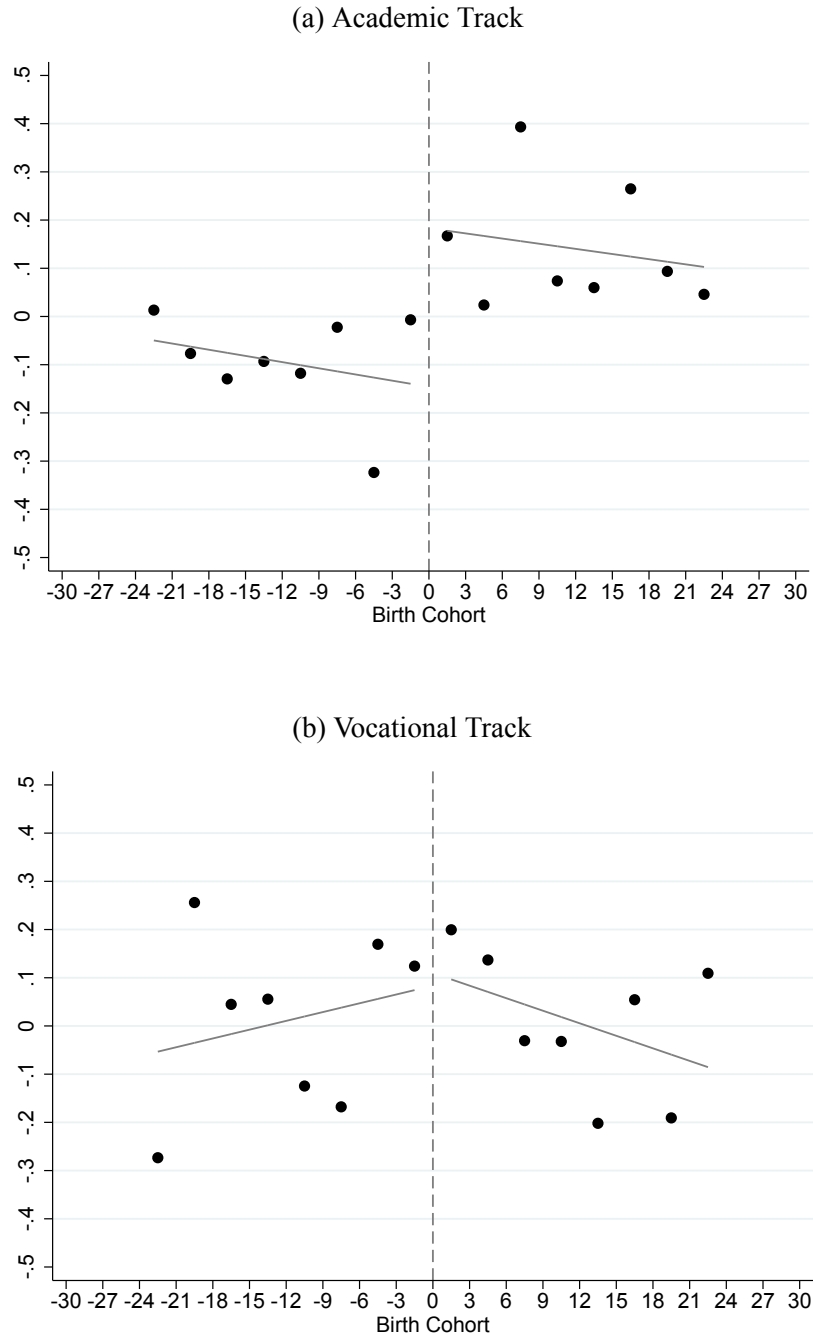


Figure 4: Taiwanese Identity and Birth Cohorts: By Education Track

Notes: We pool data from 2003, 2004, 2005 TSCS and use the sample born between September 1982 and August 1986. Figure 4a includes respondents whose final education level is senior high school or university. Figure 4b includes respondents whose educational level is junior high school, vocational high school, and vocational university. We first regress *Identity* on survey year dummies and then collapse the residuals at birth year-quarter level (i.e. three birth year-month cohorts) to derive the dots. Thus, the first dot in this figure represents average *Identity* (after controlling for the survey year fixed effect) for those born in September, October, and November 1982 and the last dot represents average *Identity* (after controlling for the survey year fixed effect) for those born in June, July, and August 1986. Fitted lines are from regression of the dots on a first order polynomial of birth year-quarter interacted with *TextBook* dummy variable.

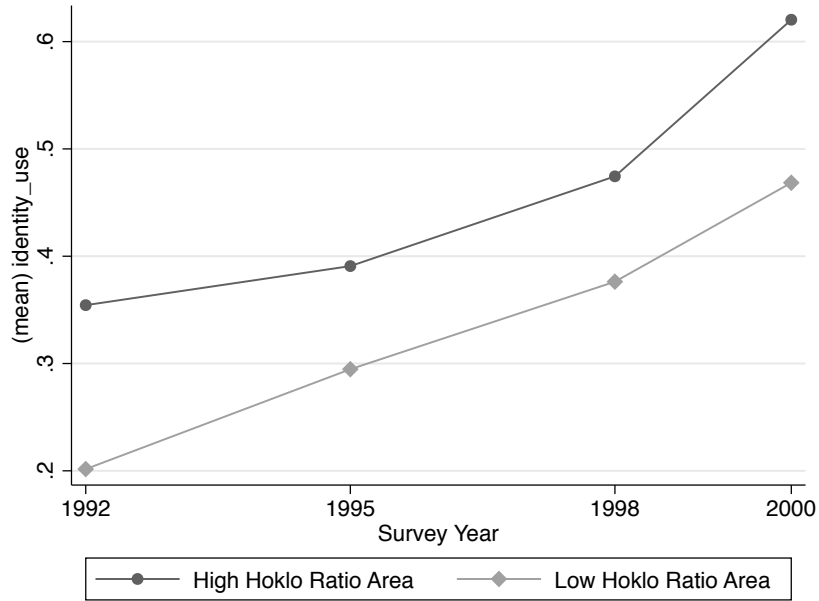


Figure 5: Taiwanese Identity Trend in High/Low *Hoklo Proportion* Areas in 1990s

Notes: We pool data from 1992, 1995, 1998 and 2000 TSCS waves. In order to include those adults whom children are more likely to meet, we drop the respondents aged below 30. We collapse the data into survey-year-*Less-Hoklo-Hometown* level. Each dot represents the mean of *Identity* in each cell.

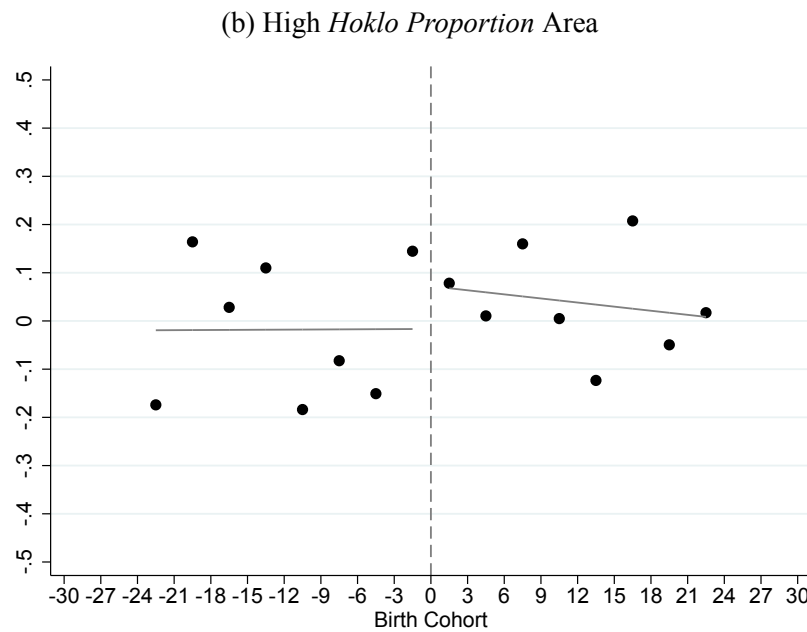
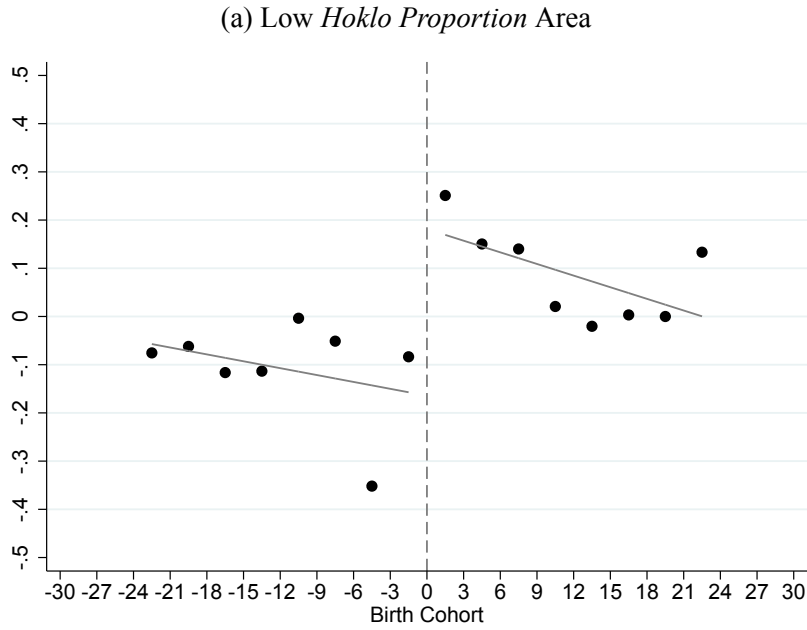


Figure 6: Taiwanese Identity and Birth Cohorts: By High/Low *Hoklo Proportion* Areas

Notes: We pool data from 2003, 2004, 2005 TSCS and use the sample born between September 1982 and August 1986. Figure 6a includes respondents responded lived longest in low *Hoklo Proportion* towns (*Less-Hoklo-Hometown* equal to 1) before age 15. Figure 6b includes respondents responded lived in high *Hoklo Proportion* towns (*Less-Hoklo-Hometown* 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. Thus, zero in the figure represents September, October, and November 1984. Fitted lines are from regression of the dots on a first order polynomial of birth year-quarter interacted with *TextBook* dummy variable.

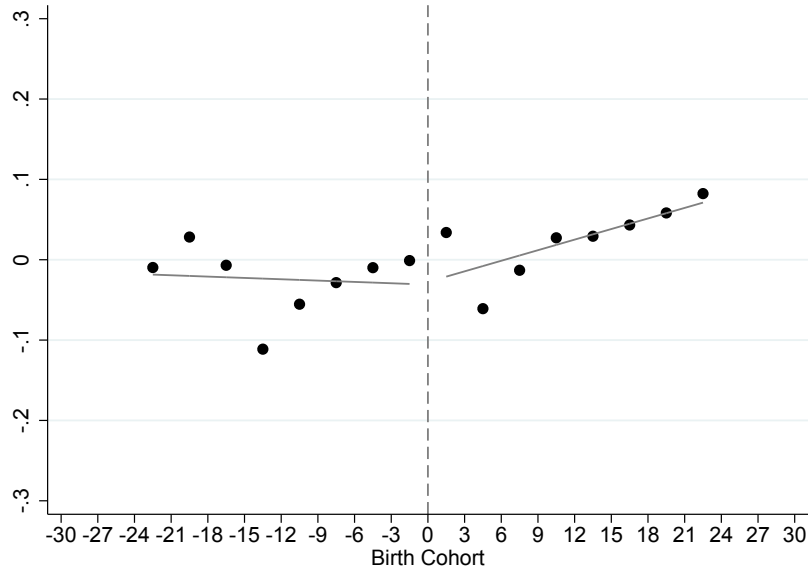


Figure 7: Taiwanese Identity and Birth Cohorts: Long-Term Results

Notes: We pool data from 2010, 2012, 2013, 2014 and 2015 TSCS waves and use the sample born between September 1982 and August 1986. We first regress *Identity* on survey year dummies and then collapse the residuals at birth year-quarter level (i.e. three birth year-month cohorts) to derive the dots. Thus, the first dot in this figure represents average *Identity* (after controlling for the survey year fixed effect) for those born in September, October, and November 1982 and the last dot represents average *Identity* (after controlling for the survey year fixed effect) for those born in June, July, and August 1986. Fitted lines are from regression of the dots on a first order polynomial of birth year-quarter interacted with *TextBook* dummy variable.

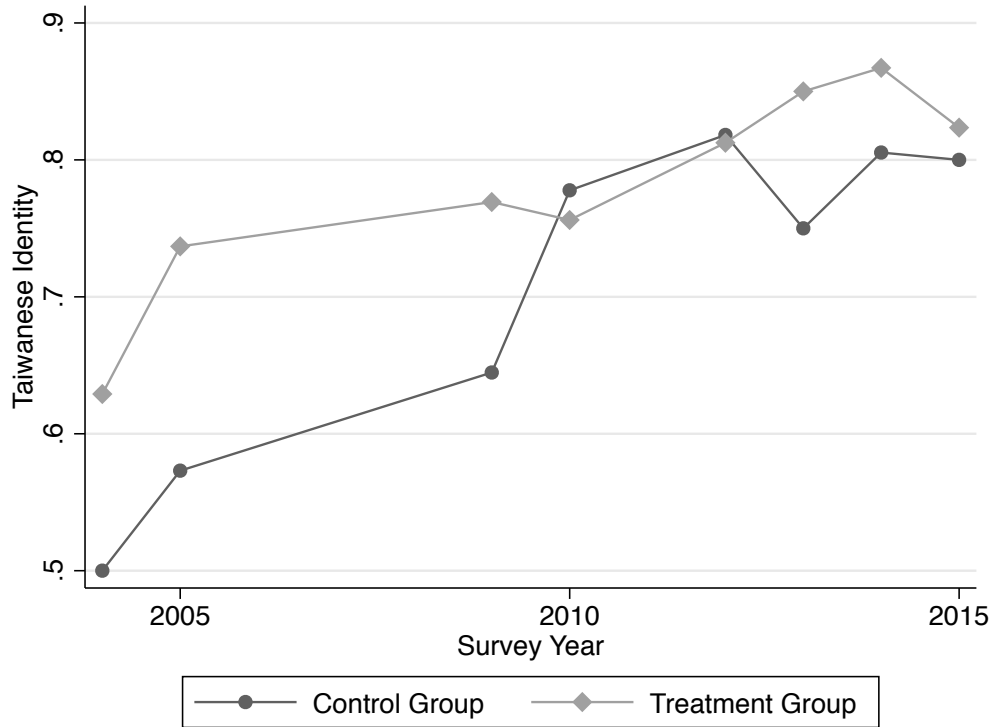


Figure 8: Trend in Taiwanese Identity: By Treatment Status

Notes: We pool data from 2004, 2005, 2009, 2010, 2012, 2013, 2014 and 2015 TSCS and use the sample born between September 1982 and August 1986. We include samples in 2009 and one of the two waves in 2014, which are not included in some of our regression analysis, since we do not need demographic variables in drawing the graph. Each dot represents 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

Tables

Table 1: The Effects of the Textbook Reform on Taiwanese Identity: Main Results

	Taiwanese Identity			
	(1)	(2)	(3)	(4)
<i>TextBook</i>	0.162** (0.0795)	0.178** (0.0796)	0.184** (0.0747)	0.176** (0.0770)
Baseline Mean	0.610	0.610	0.608	0.608
Persuasion Rate	42.6	45.7	46.9	44.8
Sample Size	427	427	417	417
R-squared	0.050	0.109	0.160	0.176
Linear Spline	Yes	Yes	Yes	Yes
Survey FE	Yes	Yes	Yes	Yes
Hometown FE	No	Yes	Yes	Yes
Demographic	No	No	Yes	Yes
Less-Hoklo-Hometown	No	No	No	Yes

Notes: We pool data from 2003, 2004, and 2005 TSCS waves and use the sample born between September 1982 and August 1986. The above table reports the coefficient of *TextBook* based on regression (1), which is 1 if the birth year-month of the respondent is after September 1984, 0 otherwise. All columns include survey year fixed effect and the first-order polynomials of birth month m interacting fully with *TextBook* (i.e. linear spline). Column (2) adds hometown fixed effect. Column (3) further includes demographic variables, such as gender, parents' education level (category variable: elementary school, junior high school, senior high school, vocational high school, college, university, military), parents ethnicity (category variable: Hoklo, Mainlanders, Hakka, Other). Column (4) adds *Less-Hoklo-Hometown* dummy variable. Baseline mean is the simple average of *Identity* of respondents born between September 1982 and August 1984. Standard errors are clustered at birth year-month level in parentheses *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 2: Robustness Check: Higher Polynomial Order

	Taiwanese Identity			
	(1)	(2)	(3)	(4)
<i>TextBook</i>	0.186 (0.107)	0.251** (0.112)	0.222** (0.102)	0.213* (0.110)
Baseline Mean	0.610	0.610	0.608	0.608
Persuasion Rate	46.9	61.9	56.6	54.3
Sample Size	427	427	417	417
R-squared	0.054	0.117	0.158	0.165
Linear Spline	Yes	Yes	Yes	Yes
Quadratic Spline	Yes	Yes	Yes	Yes
Survey FE	Yes	Yes	Yes	Yes
Hometown FE	No	Yes	Yes	Yes
Demographic	No	No	Yes	Yes
Less-Hoklo-Hometown	No	No	No	Yes

Notes: We pool data from 2003, 2004, and 2005 TSCS waves and use the sample born between September 1982 and August 1986. The above table reports the coefficient of *TextBook* based on regression (1), which is 1 if the birth year-month of the respondent is after September 1984, 0 otherwise. All columns include survey year fixed effect and the first/second-order polynomials of birth month m interacting fully with *TextBook* (i.e. linear/quadratic spline). Column (2) adds hometown fixed effect. Column (3) further includes demographic variables, such as gender, parents' education level (category variable: elementary school, junior high school, senior high school, vocational high school, college, university, military), parents ethnicity (category variable: Hoklo, Mainlanders, Hakka, Other). Column (4) adds *Less-Hoklo-Hometown* dummy variable. Baseline mean is the simple average of *Identity* of respondents born between September 1982 and August 1984. Standard errors are clustered at birth year-month level in parentheses *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 3: Robustness Check: Examine 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 Proportion
<i>TextBook</i>	0.0478 (0.0601)	0.00106 (0.0495)	0.0542 (0.0572)	-0.0703 (0.0773)	-0.0229 (0.0485)	-0.0997 (0.484)	0.0348 (0.238)	0.0331 (0.0850)	-0.122 (0.208)	0.155*** (0.0512)
Sample Size	417	417	417	417	417	417	417	417	417	417
R-squared	0.012	0.008	0.004	0.009	0.002	0.005	0.010	0.084	0.028	0.010

Notes: We pool data from 2003, 2004, and 2005 TSCS waves and use the sample born between September 1982 and August 1986. The above table reports the coefficient of *TextBook* based on regression (1), which is 1 if the birth year-month of the respondent is after September 1984, 0 otherwise. All columns include survey year fixed effect and the first-order polynomials of birth month m interacting fully with *TextBook* (i.e. linear spline). The construction of the outcome 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 proportion is the dummy variable *Less-Hoklo-Hometown*, see construction in section 3 **The proxy of hometown ethnicity distribution**. Standard errors are clustered at birth year-month level in parentheses *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 4: Robustness Check: Placebo Test of Fake Textbook Reform

	Taiwanese Identity			
	(1)	(2)	(3)	(4)
Panel A: Use September 1983 as Cut-Off				
<i>TextBook</i>	-0.0552 (0.0730)	-0.0482 (0.0685)	-0.0588 (0.0782)	-0.0449 (0.0804)
Sample Size	487	487	479	479
R-squared	0.032	0.070	0.131	0.135
Panel B: Use September 1982 as Cut-Off				
<i>TextBook</i>	-0.0085 (0.0919)	0.0100 (0.0986)	-0.0056 (0.0956)	-0.0050 (0.0994)
Sample Size	509	509	500	500
R-squared	0.030	0.066	0.131	0.131
Panel C: Use September 1981 as Cut-Off				
<i>TextBook</i>	-0.0184 (0.0686)	-0.0124 (0.0686)	0.0276 (0.0710)	0.0150 (0.0736)
Sample Size	519	519	509	509
R-squared	0.020	0.060	0.158	0.172
Linear Spline	Yes	Yes	Yes	Yes
Survey FE	Yes	Yes	Yes	Yes
Hometown FE	No	Yes	Yes	Yes
Demographic	No	No	Yes	Yes
Less-Hoklo-Hometown	No	No	No	Yes

Notes: We pool 2003, 2004, 2005 TSCS data. Panel A uses the sample born between August 1981 and September 1985; Panel B uses the sample born between August 1980 and September 1984; Panel C uses the sample born between August 1979 and September 1983. The above table reports the coefficient of *TextBook* based on regression (1). In the placebo tests, we define dummy variable *TextBook* as respondents born after September 1983, 1982 and 1981. Specifications in each column are the same as in corresponding columns in Table 1. Standard errors are clustered at birth year-month level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 5: Subgroup Analysis: By Education Track

	Taiwanese Identity			
	(1)	(2)	(3)	(4)
Panel A: Academic Track				
<i>TextBook</i>	0.279** (0.113)	0.305** (0.124)	0.283** (0.122)	0.276** (0.126)
Baseline Mean	0.583	0.583	0.580	0.580
Sample Size	219	219	214	214
R-squared	0.077	0.128	0.212	0.213
Panel B: Vocational Track				
<i>TextBook</i>	0.0546 (0.104)	0.0510 (0.103)	0.0500 (0.112)	0.0464 (0.112)
Baseline Mean	0.639	0.639	0.638	0.638
Sample Size	208	208	203	203
R-squared	0.069	0.182	0.301	0.313
Linear Spline	Yes	Yes	Yes	Yes
Survey FE	Yes	Yes	Yes	Yes
Hometown FE	No	Yes	Yes	Yes
Demographic	No	No	Yes	Yes
Less-Hoklo-Hometown	No	No	No	Yes

Notes: We pool 2003, 2004, 2005 TSCS data and use the sample born between September 1982 and August 1986. The above table reports the coefficient of *TextBook* based on regression (1), which is 1 if the birth year-month of the respondent is after September 1984, 0 otherwise. Panel A includes respondents whose final education level is senior high school or university. Panel B includes respondents whose educational level is junior high school, vocational high school, and vocational university. Specifications in each column are the same as in corresponding columns in Table 1. Baseline mean is the simple average of *Identity* of respondents born between September 1982 and August 1984 in the corresponding subgroup. Standard errors clustered at birth year-month level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 6: Subgroup Analysis: By Hometown Ethnicity Distribution

	Taiwanese Identity			
	(1)	(2)	(3)	(4)
Panel A: Hometown with Low Hoklo Proportion				
<i>TextBook</i>	0.328*** (0.113)	0.341*** (0.120)	0.366*** (0.125)	0.366*** (0.125)
Baseline Mean	0.558	0.558	0.555	0.555
Sample Size	193	193	186	186
R-squared	0.113	0.147	0.222	0.222
Panel B: Hometown with High Hoklo Proportion				
<i>TextBook</i>	0.0239 (0.103)	0.0208 (0.100)	0.0321 (0.0980)	0.0321 (0.0980)
Baseline Mean	0.650	0.650	0.647	0.647
Sample Size	234	234	231	231
R-squared	0.023	0.139	0.213	0.213
Linear Spline	Yes	Yes	Yes	Yes
Survey FE	Yes	Yes	Yes	Yes
Hometown FE	No	Yes	Yes	Yes
Demographic	No	No	Yes	Yes
Less-Hoklo-Hometown	No	No	No	Yes

Notes: We pool 2003, 2004, 2005 TSCS data and use the sample born between September 1982 and August 1986. The above table reports the coefficient of *TextBook* based on regression (1), which is 1 if the birth year-month of the respondent is after September 1984, 0 otherwise. Panel A includes respondents whose hometown has lower Hoklo proportion comparing to the median of the population in National Hakka Population Basic Information Survey Research, while Panel B includes respondents whose hometown has higher ratio. Specifications in each column are the same as in corresponding columns in Table 1. Baseline mean is the simple average of *Identity* of respondents born between September 1982 and August 1984 in the corresponding subgroup. Standard errors clustered at birth year-month level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 7: The Effects of the Textbook Reform on Taiwanese Identity: Long-Term Results

	Taiwanese Identity			
	(1)	(2)	(3)	(4)
<i>TextBook</i>	0.0089 (0.0497)	-0.0021 (0.0499)	-0.0269 (0.0583)	-0.0252 (0.0589)
Baseline Mean	0.771	0.771	0.784	0.784
Sample Size	963	963	655	655
R-squared	0.018	0.058	0.152	0.153
Linear Spline	Yes	Yes	Yes	Yes
Survey FE	Yes	Yes	Yes	Yes
Hometown FE	No	Yes	Yes	Yes
Demographic	No	No	Yes	Yes
Less-Hoklo-Hometown	No	No	No	Yes

Notes: We pool the 2009, 2010, 2012, 2013, 2014 and 2015 TSCS waves and use the sample born between September 1982 and August 1986. The above table reports the coefficient of *TextBook* based on regression (1), which is 1 if the birth year-month of the respondent is after September 1984, 0 otherwise. Specifications are the same as Table 1. Baseline mean is the simple average of *Identity* of respondents born between September 1982 and August 1984. Clustered standard errors at birth year-month level in parentheses.*** p<0.01, ** p<0.05, * p<0.1

Online Appendix

A Hoklo Ethnicity Distribution in Taiwan

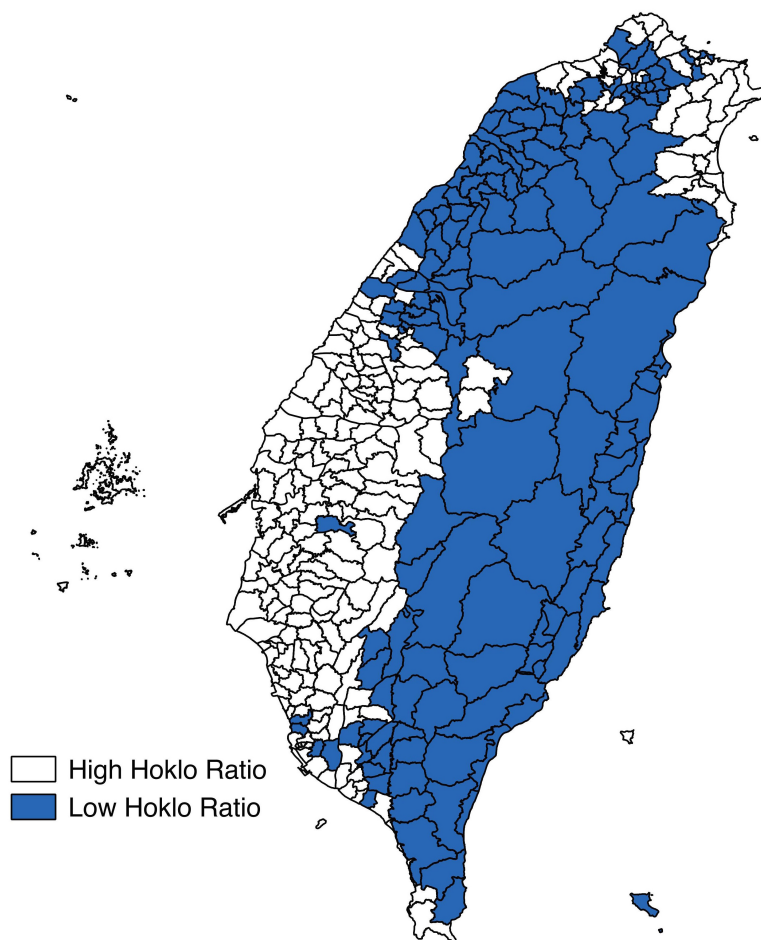


Figure A1: The Map with High/Low Hoklo Proportion at Town Level

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 Proportion* towns has proportion of Hoklo people higher than 77.1%; otherwise, the town belongs to low *Hoklo Proportion* towns.