WORKING PAPER SERIES

Chinese Language Learning Opportunities Abroad and International Student Mobility to China

Yue Huang/Jingjing Lyu

Working Paper No. 8/2019

OTTO VON GUERICKE
UNIVERSITÄT
MAGDEBURG

FACULTY OF ECONOMICS
AND MANAGEMENT

Impressum (§ 5 TMG)

Herausgeber:
Otto-von-Guericke-Universität Magdeburg
Fakultät für Wirtschaftswissenschaft
Der Dekan

Verantwortlich für diese Ausgabe:
Yue Huang and Jingjing Lyu
Otto-von-Guericke-Universität Magdeburg
Fakultät für Wirtschaftswissenschaft
Postfach 4120
39016 Magdeburg
Germany

http://www.fww.ovgu.de/femm

Bezug über den Herausgeber
ISSN 1615-4274

Chinese Language Learning Opportunities Abroad and International Student Mobility to China

Yue Huang
Otto von Guericke University Magdeburg.
Email: yue.huang@ovgu.de

Jingjing Lyu[†]
Westfälische Wilhelms-Universität Münster, Department of Economics (CQE),
Am Stadtgraben 9, 48143 Münster, Germany.
Email: jlyu@wiwi.uni-muenster.de. Phone: +49 176 2407 0642.

Abstract

The number of overseas students in China has increased substantially over the last two decades, as has the number of Confucius Institutes (CIs) abroad. Using both official and self-compiled data on CIs abroad, and overseas students in China, by country of origin, we investigate empirically whether Chinese language learning opportunities abroad have exerted a positive effect on the number of students who move to China for study. Using panel data for 182 countries over the period 2002 to 2014, we find evidence in fixed-effects regressions for a sizeable positive effect of the number of CIs in a country on the number of overseas students from that country in China. We also find evidence for effect heterogeneity by countries' geographic distance to China, the linguistic distance from their official language to Chinese, and their income levels.

Keywords: Confucius Institute, Overseas Student, Migration, China
JEL Classification: F22, I20, N35.

Acknowledgement: This paper has benefited from comments by Michael Kvasnicka, Bernd Süssmuth and participants of the Annual Conference of the European Society for Population Economics in Bath, the 4th Bamberg-Halle-Jena-Leipzig Workshop on Empirical Microeconomics / Applied Microeconometrics, and seminars at Otto von Guericke University Magdeburg, the University of Leipzig and RWI - Leibniz Institute for Economic Research. All remaining errors are our own.

[†]Corresponding author.

1 Introduction

In recent years, a fast growing body of literature has emerged on foreign language learning opportunities, language institutes, and their economic impacts.[1] Many studies focus on the effects of language learning opportunities in the home country of foreign learners, as measured by the number of language institutes, on trade and foreign direct investment (FDI) between the country operating the institutes and the countries hosting them (Akhtaruzzaman et al., 2017; Ghosh et al., 2017; Lien et al., 2012; Lien and Co, 2013; Lien and Lo, 2017; Lien et al., 2019). The majority of these studies find a positive association between the number of language institutes and economic outcomes. However, they failed to explore the reasons for such links. In this study, which focuses on China and its Confucius Institutes (CIs), we explore one potential reason, namely the growth in host-country student numbers in the country that operates language institutes abroad, by studying empirically whether the opening of CIs has a positive effect on the number of overseas students in China.

With the rapid industrial and economic growth that is occurring in China, more and more people from all over the world have become interested in traveling to China for various purposes, such as tourism, work, and study. In 2002, around 85,000 overseas students studied in China for periods ranging from short to long[2] (Department of Policy Planning of the Ministry of Foreign Affairs of the People's Republic of China, 2003), and this number had surged to 492,185[3] by 2018. Apart from China's rapid and steady economic growth, Chinese culture is also a driving force for the significant increase in the number of overseas students China hosts. Learning the Chinese language, especially in the home country of the learners, not only provides them with language skills, but also helps foreigners to obtain more information about China, both of which can encourage them to travel to China. With the mission to promote Chinese language learning and teaching, and to conduct and provide information on language and cultural exchange activities between China and other countries, the *Confucius Institute Headquarters*, also known as Hanban, founded the first CI in South Korea in 2004. Since then, the number of CIs and their registered students have increased exponentially. By the end of 2018, there were 548 CIs in 147 countries (and regions, i.e. Hong Kong and Macao) (Confucius Institute Headquarters (Hanban), 2018). In the recent

[1] Studies in this literature include Akhtaruzzaman et al. (2017), Ghosh et al. (2017), Lien et al. (2012), Lien and Co (2013), Lien et al. (2014), Lien and Oh (2014), Lien et al. (2017), Lien and Lo (2017), Lien et al. (2019), Lin et al. (2016), and Huber and Uebelmesser (2019).

[2] According to the *Ministry of Education of the People's Republic of China*, students who study in China for six months or less are defined as short-term students, and those who study in China for more than six months as long-term students. For more information, see: http://www.moe.gov.cn/s78/A20/gjs_left/moe_850/tnull_8292.html.

[3] This figure is provided by the *Ministry of Education of the People's Republic of China* at: http://www.moe.gov.cn/jyb_xwfb/gzdt_gzdt/s5987/201904/t20190412_377692.html.

past, therefore, the numbers both of overseas students in China and of CIs abroad, have increased dramatically and in tandem.

Chinese is the language with the most native speakers worldwide. Following English, Chinese is also the second most frequently spoken language if only second-language speakers are counted (Grenier, 2015). As an investment in human capital, learning a dominant or common language has a proven economic value (Oh et al., 2011). If members of society can communicate with each other using the same language, economic well-being can be enhanced (Grenier, 2015), since communication and the exchange of information are of great importance for improving understanding various economic, social, and cultural issues (Čok and Novak-Lukanovič, 2005). However, learning a new language is not cost-free. Mastering a new language takes substantial time, money, and resources that could be applied to other activities (Grenier, 2015). Opening language institutes and providing Chinese learning opportunities in the home country of foreign learners can help to reduce some of these costs, which can then stimulate more foreigners to learn Chinese.

Higher education has become a global business and in the last few decades, international student mobility has continued to rise (Bessey, 2012). Overseas students constitute one type of international migration that has become an integral feature of an increasingly interconnected world. Students who have experience in education abroad tend to demonstrate linguistic improvement, positive cultural experience from living in another country, general personal development and better career prospects (King and Ruiz-Gelices, 2003). Learning the language of the destination country, especially before traveling there, is beneficial for the quality of students' study and life in that country. Given that most courses at a Chinese university are conducted in Chinese, overseas students may be required to master a certain level of the language and even provide a language test certificate when they submit their applications. The literature on general migration and language acquisition also indicates that proficiency in the language of the destination country can improve labor market outcomes, such as earnings and employment (Chiswick and Miller, 1995; Dustmann and van Soest, 2001; Berman et al., 2003; Dustmann and Fabbri, 2003; Hwang et al., 2010). Therefore, the acquisition of the destination-country language is not only essential for overseas study, but also connected to the social rules and local culture that are the main factors for social and economic integration within the society, which is of great importance for overseas students in addition to their study.

Since the establishment of CIs, China has greatly enhanced its soft power[4] on the global stage (Nye, 2005). A strand of the literature qualitatively discusses the origin and effects

[4]Soft power is defined as the capability of a country to get other countries to *want* what it wants through "soft" means, such as cultural attraction or international institutions (Nye, 1990).

of soft power and its relationship with the official language learning institutes founded by a country (Groot, 2018; Lo and Pan, 2016; Nye, 2005, 2017). Since the Chinese government proposed its economic reforms policy in 1978, China has experienced decades of rapid economic growth and is keen to export not only manufactured goods, but also its culture and history. With the CI project, China has expanded its soft power diplomacy to brand the nation and promote harmonious international relationships (Lo and Pan, 2016). Another strand of literature has argued that language institutes are a tool for reducing transaction costs, which in turn increases international trade and communication (Akhtaruzzaman et al., 2017; Ghosh et al., 2017; Lien and Co, 2013; Lien et al., 2012, 2014, 2017, 2019; Oh et al., 2011), and a means of motivating international migration to the institute-operating country (Huber and Uebelmesser, 2019). However, little research has been conducted on the relationship between (Chinese) language learning opportunities and overseas students. Using panel data of 18 countries that have the most students studying in China from 2003 to 2012, Miao and Chen (2015) demonstrate positive effects of the number of CIs on the number of overseas students in China. They also find effect heterogeneity by country trade volumes with China. Using panel data for 40 countries from 2004 to 2014, Lin et al. (2016) investigate the effects of CIs on the number of overseas students in China. The study finds, surprisingly and counterintuitively, that the establishment of a CI causes a decrease in the number of overseas students in China by 0.3%. The study also finds evidence for effect heterogeneity by countries' income level. Lien et al. (2018), using a relatively larger sample of 53 countries, provide evidence for a positive association between CIs and the number of overseas students in China (termed China's education exports). They also find effect heterogeneity by cultural differences between China and CI-host countries and the institutional quality in CI-host countries. Using provincial data on Chinese partner universities of CIs abroad and foreign students in China for the period 2004 – 2015, Lien and Miao (2018) find a positive association between the number of Chinese partner universities and non-degree foreign-student numbers in China. However, none of the studies on CIs and overseas students explicitly discuss the potential endogeneity of CI establishment.

In this paper, we investigate the effects of CIs abroad on the number of overseas students in China in a fixed-effects regression framework for an unbalanced panel data set of 182 countries in the period 2002 to 2014. Our study contributes to the literature on soft power and language institutes and adds to the small but growing body of empirical studies on foreign language learning opportunities. In particular, our study is the first to provide a comprehensive analysis of the effects of CIs on overseas students in China. Using more elaborate CI data and considering a total of 182 countries over 13 years, we analyze a much larger estimation sample than Lin et al. (2016), who study only a selective sample of 40

countries which host the most overseas students, a feature that renders the study and its counterintuitive findings susceptible to sample selection bias. Our results show, in contrast to those of Lin et al. (2016), a sizeable positive effect of the number of CIs abroad on the number of overseas students in China. We also test much more thoroughly than most of the literature, for potential confounders and factors that may induce endogeneity of language learning opportunities. Specifically, we explicitly consider various factors that may determine the establishment of a CI, examine changes in the number of overseas students before the opening of CIs, and also control for region-specific trends. Our main findings prove robust in their sensitivity and causality checks. We also find that establishing CIs has a larger effect on overseas student numbers in countries that are geographically closer to China, linguistically closer to Chinese, and poorer in terms of average income levels.

The remainder of this paper is structured as follows. Section 2 provides a brief introduction to CIs and discusses the literature on language institutes and their impact. Section 3 describes the data we employ, defines all variables we use in the empirical analysis, and outlines our estimation strategy. Section 4 presents our main results, reports several robustness checks, and provides analyses that explore potential effect heterogeneity. Finally, Section 5 summarizes our main findings and concludes.

2 Background

In this section, we first provide a brief introduction to Confucius Institutes (CIs) abroad and overseas students in China, present the regional (country-level) distribution of CIs and of overseas students in China by country of origin, and then document their evolution in recent decades (Section 2.1). After that, in Section 2.2, we summarize the findings of empirical studies on language institutes, and review solutions for dealing with the potential endogeneity of language institutes discussed in the literature.

2.1 Confucius Institutes and overseas students in China

As China's economic and cultural exchange with the world has risen rapidly, worldwide demand for learning about the Chinese culture and language has increased sharply. In 2004, China began launching non-profit public educational organizations (CIs) in other countries to promote global knowledge and awareness of the Chinese language and culture, to support Chinese teaching internationally, to facilitate cultural and educational exchange between China and other countries, and to enhance more generally the understanding between Chi-

nese people and those from the rest of the world.[5] CIs are initiated by *Confucius Institute Headquarters* (Hanban), a public institution affiliated with the *Chinese Ministry of Education*. The main functions of Hanban include the following aspects: (i) to support educational institutions abroad with respect to teaching Chinese and disseminating Chinese culture; (ii) to compile, improve and promote the *Standard for International Chinese Language Teachers*, *International Standard on Chinese Language Proficiency* and *General Outline of the International Chinese Language Teaching*; (iii) to select and qualify Chinese teachers; (iv) to organize the *Chinese Proficiency Test* (HSK); and (v) to construct an international platform for Chinese teaching through networking, television, and radio and to provide digital resources for learning and teaching Chinese.[6] A CI is hosted by a foreign partner organization, co-operated by a Chinese institution, and owned by Hanban.[7] By the end of 2018, there were 548 CIs in 147 countries (and regions) (Confucius Institute Headquarters (Hanban), 2018). The United States has the most worldwide. Other countries with a large number of institutes (more than 10) include, for example, South Korea, Japan, Thailand, Germany, Russia, France, Italy, the United Kingdom, Canada, and Australia.

With the expansion of CIs abroad, the number of overseas students traveling to China has also increased significantly. Chinese is now popular as a study major among overseas students in the manner of medicine, engineering, economics, and management. On the one hand, foreign students need to master a sufficient amount of the Chinese language to be enrolled in some Chinese universities, and in general, better language skills of course make life in China easier. Therefore, it is perceived as important and beneficial for students to learn Chinese. Opening CIs in students' home country provides them with the chance to learn Chinese before they visit China for study. On the other hand, because of the increasing establishment of CIs abroad, it is much easier for foreigners to have opportunities to learn Chinese and become familiar with Chinese culture, which may prompt them to migrate to China for study.

In Figure 1, we standardize the number of CIs[8] and students in 2014 by 1 million pop-

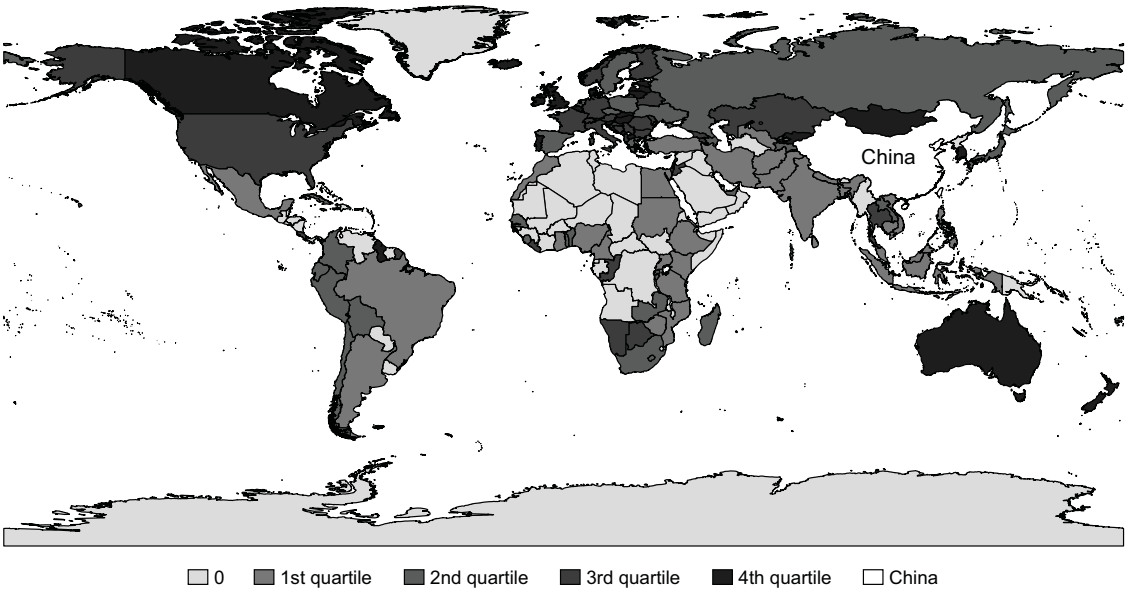
[5]More information about the functions of CIs can be found at: http://english.hanban.org/node_7716.htm.

[6]More information about the functions of Hanban can be found at: http://english.hanban.org/node_7719.htm.

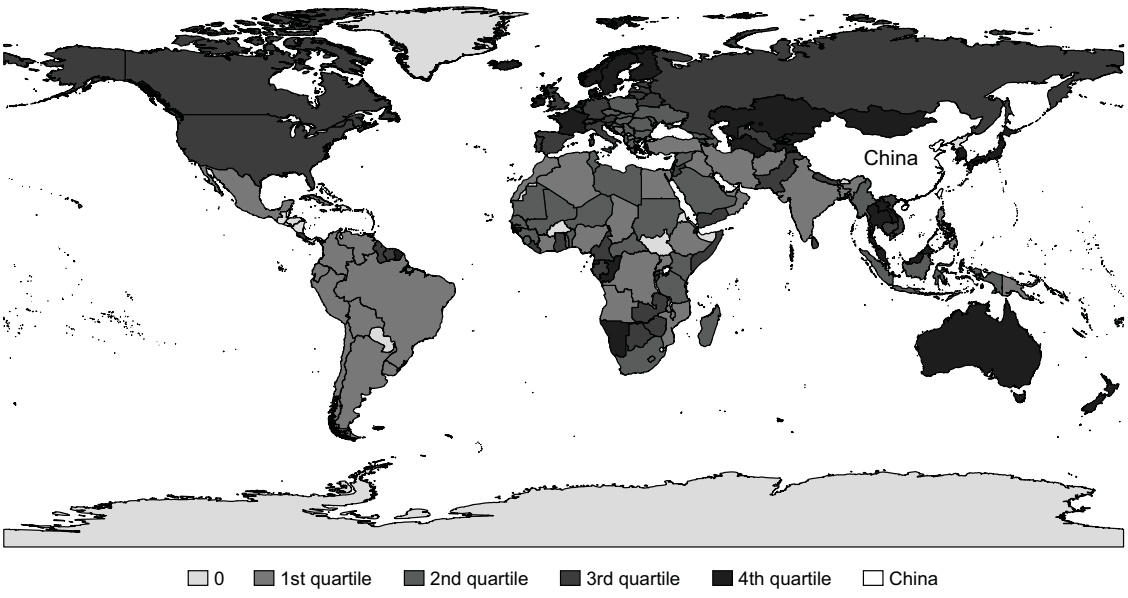
[7]Another organizational form of Hanban to provide Chinese language learning opportunities abroad is "Confucius Classrooms". A Confucius Institute is established as an "executive institution" in a foreign country with the expectation of ongoing and lasting operation, whereas a Confucius Classroom is usually built into a previously existing institution, for instance a university-sponsored institution, and may operate for only a short period of time (Lien et al., 2014). Since the exact running times of all classrooms is not fully available, in this study we focus only on Confucius Institutes.

[8]Only institutes that are already launched are counted. Those in planning or preparing process, e.g. with a signed letter of intent or agreement, are not counted. This also applies to Figure 2. Detailed information on CI data is provided in Section 3.1.

FIGURE 1: The number of Confucius Institutes abroad and overseas students in China in 2014 per 1 million population aged 15 – 34



(a) population standardized number of CIs

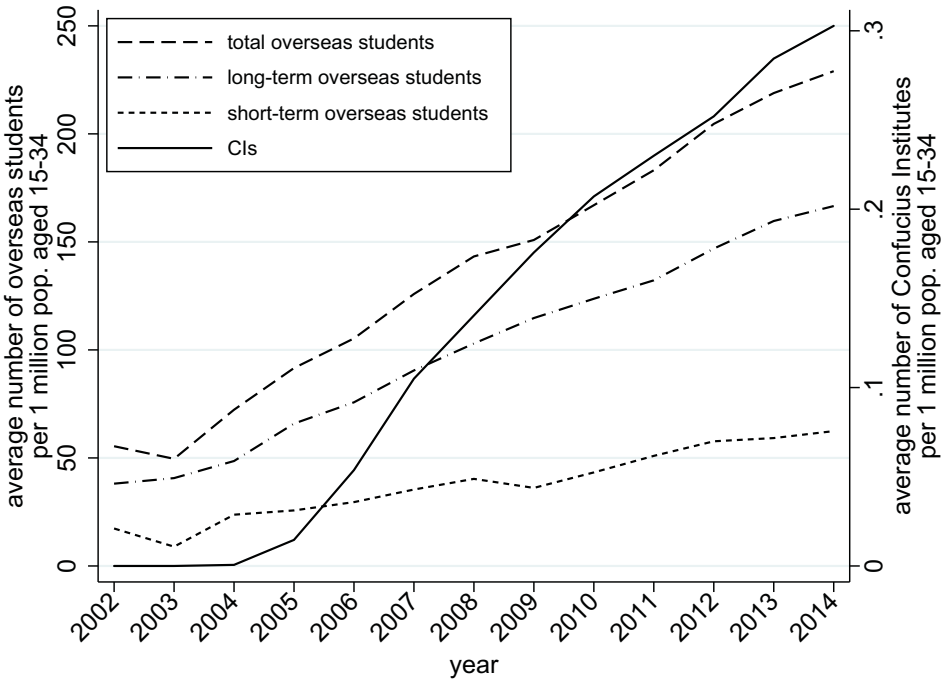


(b) population standardized number of overseas students

Data source: Hanban, China's Foreign Affairs, World Bank, own calculation.
Notes: Panel (a) shows the number of CIs per 1 million population aged 15 – 34 in 2014, and Panel (b) the number of overseas students in China per 1 million population aged 15 – 34 of origin countries in 2014.

ulation aged 15 – 34 of the host/origin country. All countries with a positive number of institutes or students are classified into four quartiles. It is evident that countries with a larger number of institutes per capita tend to send more students to China. Figure 2 plots the average number of overseas students in China per country and the number of CIs abroad per annum in the years 2002 to 2014.[9] As can be seen, the average number of overseas students and CIs are positively correlated and increased steadily over time. Whether this positive correlation reflects a causal relationship, however, is unclear. In this paper, we explore this link, i.e. whether CIs, a measure of Chinese learning opportunities abroad, exert a positive causal effect on the number of overseas students in China.

FIGURE 2: Overseas students in China and Confucius Institutes abroad, 2002 – 2014



Notes: This figure shows the evolution of CIs and overseas students from 2002 to 2014. The data on overseas students are extracted from China's Foreign Affairs (2003 – 2015), and data on Confucius Institutes are provided by Hanban. Only institutes already in operation are counted. 182 countries are observed. Both the number of overseas students and Confucius Institutes are standardized by 1 million population aged 15 – 34 of a country and weighted by a country's population in 2002.

[9]We employed our later estimation sample to plot this figure, which covers 182 countries. The average number of students and CIs per country were obtained by using countries' 2002 populations as weights. The population share of CIs can be very large if a country has a small population. For those countries, a smaller weight was assigned. Information on the data, variables, and estimation sample are discussed in detail in Section 3. In the estimation sample, some countries are dropped due to missing values in the control variables. If we deal with all countries with complete information on the number of CIs and overseas students to plot their development over time, without considering other covariates, similar patterns can be found in which both the number of CIs and overseas students increased from 2002 to 2014 (not shown).

2.2 Literature review

The strand of literature on overseas students mainly investigates their performance and the determinants of their mobility. Roy et al. (2019) provide a comprehensive review of literature that studies the cultural, personal, and employment/career outcomes of students who participate in short-term international mobility programs, and conclude that participation in such programs has a positive effect on students' cultural awareness, global mindedness, language proficiency, academic performance, professional development, perceived employability, career success and so on. Using a panel data set of bilateral flows of Erasmus students for all the participating countries, González et al. (2011) show that population in both host and home country, distance between host and home country, cost of living, university quality, language and climate in the host country are significant determinants on students' mobility. Findlay (2011) points out that marketing the opportunities to study in the UK by British universities and organizations such as the British Council has been effective in recruiting international students. Collins (2008) argues that education agents, immigrant entrepreneurs and interpersonal relationships facilitate the movement of international students from South Korea to Auckland.

The strand of literature investigating the effects of language learning opportunities abroad, especially the impact of CIs, has been increasing rapidly in recent years. Employing gravity models, Lien et al. (2012) investigate CIs' impact on outward trade and FDI flows from China, and find that CIs contribute positively to the volume of exports and FDI flows from China to other developing countries. Similarly, Lien and Co (2013) report a 5%–6% increase in state-level exports from the US to China for each additional CI established in the US between 2006 and 2010, and Akhtaruzzaman et al. (2017) provide evidence that CIs had a positive effect on Chinese FDI outflows to African countries from 2004 to 2012. Using data on tourism flows from other countries to China between 2004 and 2010, Lien et al. (2014) find a positive effect of the number of CIs on overall tourism, and on business in addition to worker tourists in particular. Lien et al. (2017) corroborate the previous findings on the effect of CIs on international travel to China, and additionally show that such effects vary with the institutional quality of host countries of CIs, as measured by indicators for economic freedom, and with cultural differences between China and host countries. Similarly, Ghosh et al. (2017) find that the presence of CIs increases tourism and equity flows to China, and exports and FDI flows from China.

Studies on the effects of language and cultural institutes from other countries have also been growing rapidly. Lien and Lo (2017) find a significant positive effect of the number of Goethe-Institutes on trade and FDI between Germany and partner countries, and of the number of British Council institutes on UK trade. Lien et al. (2019), who analyze the effects

of three language and cultural institutes, CIs, Goethe-Institutes and Cervantes Institutes, corroborate early findings of a positive effect of the number of institutes on trade and FDI outflows from the institute-operating country (China, Germany and Spain, respectively) to other countries. They also show that this positive effect is stronger if host countries have developing economies and that the effect on FDI is larger than that on trade. Recent work by Huber and Uebelmesser (2019) finds a positive association between the number of Goethe-Institutes in a country and migration from that country to Germany.

For outcomes other than international trade, investment, tourism, and (general) migration, the literature on language institutes and overseas students is scarce. Most relevant for our paper are the empirical studies of Miao and Chen (2015), Lin et al. (2016), and Lien et al. (2018). Miao and Chen (2015) use a small sample of only 18 countries with the most overseas students, and study the relationship between overseas education consumption and the number of CIs abroad. They find a positive effect of CIs on overseas education consumption; the effect is stronger for countries with more trade with China. Lin et al. (2016) study overseas students in China between 2004 and 2014, but find only a positive correlation between CIs and overseas students for developing countries, and a negative association for all observed countries. Since both Miao and Chen (2015) and Lin et al. (2016) only use small samples of selected high overseas-student-intensive countries, their findings are likely to suffer from sample selection bias, limiting the representativeness and explanatory power of their analysis. Similarly, Lien et al. (2018) use a sample of 53 countries for the period 2000 – 2014, which still falls significantly short of the number of countries with CIs in this period. This study finds a positive association between the number of CIs abroad and the number of overseas students in China. However, the authors fail to control for country fixed effects, which may introduce bias if time-invariant country features are correlated with the establishment of CIs and the number of overseas students from a particular country in China.

Language institutes are unlikely to be randomly located across countries. Lien and Oh (2014), who analyze the determinants of CI establishments, produce evidence which suggests that international trade and distance between China and potential host countries, in addition to host country population, economic conditions and the use of English as a major spoken language, are important determinants of the opening of a CI. Empirical studies usually control for these factors by accounting for country-fixed effects and using appropriate covariates in their regression analysis. However, it is still possible that some (un-)observed and uncontrolled factors can influence the number of established CIs and outcomes of interest. For instance, to enhance educational exchange with a country, China may provide more scholarships to overseas students from a particular country and also open new CIs there. Moreover, such potentially confounding behavior may vary not only across countries

but also over time, which makes it difficult to identify the causal effect of CIs on overseas student numbers. The literature rarely explicitly discusses such issues of endogeneity. Lien et al. (2012, 2017, 2018); Lien and Lo (2017) and Lin et al. (2016) try to reduce or mitigate the endogeneity problem by measuring (taking stock of) the number of CIs one year prior to the measurement of outcomes. However, (un-)observable factors can still be correlated with both outcomes and CIs, even they are measured in different periods. Therefore, a lagged measure of an independent variable may not solve the endogeneity problem in a panel data set. Lien et al. (2012) employ the lagged dependent variable as an explanatory variable and use the system-generalized method of moments (system GMM) to estimate the effects of CI numbers on trade and FDI flows. System GMM allows variation in the outcome variable in early periods to cause new openings of CIs. However, system GMM requires that the difference of the outcome variable between two periods is not correlated with the individual (country) fixed effect[10] (Blundell and Bond, 1998; Chen, 2014), which is hard to test and usually ignored altogether in empirical studies. System GMM can be applied if countries are not too far from steady states (Chen, 2014; Roodman, 2009). Neither of these two conditions are likely to be fulfilled for all countries and in all years. What is more, unobserved factors contained in error terms could still be correlated with CI numbers. Hausman-Taylor instrumental variable estimation is sometimes used in the literature to address the endogeneity problem (Lien et al., 2012). However, there are general concerns about the validity of instrumental variables used (Lien et al., 2017, 2019). Huber and Uebelmesser (2019) provide evidence that the establishment of Goethe-Institutes affects migration flows to the German-speaking regions of Switzerland, but not to the French- and Italian-speaking regions, suggesting that the German-speaking regions of Switzerland benefit from German language institutes, whose establishment, however, is assumed to be uncorrelated with other determinants of migration flows to these regions of Switzerland.

In this study, we use a much larger sample of countries than the existing literature on China, and study the impact of CIs on overseas students in China. We also discuss and address the endogeneity problem more thoroughly than the literature, which provides (greater) credence for a causal interpretation of our findings.

3 Data and Empirical Strategy

We use three main data sources to analyze the effect of CIs abroad on the number of overseas students in China. The first data extract consists of annual statistics on country-level overseas students in China from 2002 to 2014 and contains information on their numbers

[10] Another assumption is that error terms have no auto-correlation (Blundell and Bond, 1998).

by type of stay (long-term or short-term study) and country of origin. The second data source contains information from 2002 to 2014, on the number of CIs abroad by country, and the third provides information on demographic and economic characteristics of overseas students' countries of origin which could influence the outflow of students to China and the establishment of a CI. Below, we first describe in greater detail these data and the variables we construct from them (Section 3.1). Thereafter, in Section 3.2, we present and discuss our empirical strategy.

3.1 Data and variables

Annual statistics on overseas students in China from 2002 to 2014 are extracted from *China's Foreign Affairs* (2003 – 2015) and record the number of overseas students who study in China in a certain year. The data contain information on: (i) the total number of overseas students in China and their numbers by country of origin; (ii) the number of overseas students in China for long-term study and their numbers by country of origin; and (iii) the number of overseas students in China for short-term study and their numbers by country of origin. Since Chinese learning opportunities abroad may have different effects on long- and short-term study in China, we consider three measures of overseas students by country as outcomes, counts of those on a long-term stay, on a short-term stay, and the summary total of both.

The second data source contains information on the number of CIs in a country, which provides a measure of Chinese learning opportunities abroad. We self-compiled these data from the official web page of Hanban.[11] Hanban records information on the location of a CI and when it opened.[12] At the country level, data extracts contain information on the total number of institutes in a country at the end of a year in the sampling period 2002 to 2014. 116 countries had at least one institute in 2014, of which the US hosted 107, followed by the UK with 25 institutes and South Korea with 20.

The last type of data we use in the analysis contains information from the World Bank and the National Bureau of Statistics of China on the demographic and economic conditions of origin countries of overseas students, such as their gross domestic product (GDP) per capita, their total population, their population aged 15 – 34, and their exports to and imports from China.

[11]Data source: http://www.hanban.org/confuciousinstitutes/, accessed in May 2017. Very few institutes closed before we accessed the data are not recorded on the web page of Hanban. Countries in which a CI was closed include the US, Germany, France, Sweden, and Canada, which are countries with a relatively large number of CIs. Appendix B provides detailed information on our collection of CI data and a brief summary of closures of CIs.

[12]The decision to open a CI could precede the actual opening significantly. Detailed information can be found in Appendix B.

Using our data, we are able to construct the dependent and main independent variables and generate other control variables. These variables are described below.

Dependent variables: *number of overseas students (in total, for long-, and for short-term study) in China per 1,000,000 population aged 15 – 34 of the country of origin.* We only exploit population aged between 15 and 34 to calculate the population share of overseas students, because they are more likely than people in other age groups to migrate to China for study.

Main independent variable: *number of CIs per 1,000,000 population aged 15 – 34 of the host country.* The main independent variable in our analysis is an indicator for Chinese learning opportunities abroad, i.e. the number of CIs in a country. It is a stock measure taken on the last day per annum in the period 2002 to 2014. We also standardize this measure per 1,000,000 population of the host country in the age group of 15- to 34-year olds.

Other covariates. In addition to our main explanatory variable, we control for potential confounders that may influence both the number of overseas students in China and the establishment of CIs abroad. We consider two types of control variables. First, we employ the logarithm of GDP per capita (in US dollars) of the country of origin of overseas students as an indicator for its economic condition. Second, we use the logarithms of exports to and imports from China as indicators for the importance of China as a trade partner for a country.

3.2 Empirical strategy

We estimate the following fixed effects model to study the effects of Chinese learning opportunities abroad, as measured by the number of CIs per 1 million capita, on the number of overseas students in China:

$$ST_{it} = \alpha_0 + \alpha_1 \, CI_{it} + \boldsymbol{\alpha_2} \, \boldsymbol{X_{i,t-1}} + \boldsymbol{d_i} + \boldsymbol{d_t} + \varepsilon_{it}, \tag{1}$$

where ST_{it} is the number of overseas students from country i who study in China in year t per 1,000,000 population aged 15 – 34 of the country of origin i in year t, which ranges from 2002 to 2014. We consider three outcome variables, the number of overseas students who go to China for long- or short-term study (ST^{long} and ST^{short}, respectively), and the sum of these two (ST^{total}). CI_{it} is the number of CIs in country i at the end of year t standardized per 1,000,000 population aged 15 – 34 of that country in the same year. $\boldsymbol{X_{i,t-1}}$ is a vector of control variables for year $t-1$ characteristics of country i, and includes the logarithm of GDP per capita (in US dollars) of country i ($lgdp_{i,t-1}$), the logarithm of the absolute value of exports to China ($lexp_{i,t-1}$), and the logarithm of the absolute value of imports from

China ($limp_{i,t-1}$). $\boldsymbol{d_i}$ controls for country fixed effects, i.e. level differences in the number of overseas students between countries caused by country-specific time-invariant factors, such as the geographic distance between country i and China. $\boldsymbol{d_t}$ controls for year fixed effects, i.e. aggregate time trends in the number of overseas students that do not differ across countries. Finally, ε_{it} is an error term. The key parameter of interest is α_1. If more Chinese learning opportunities abroad increase the number of students who move to China to study, α_1 should be positive. Observations are weighted by a country's population in the year 2002, which is prior to the establishment of CIs, so that the estimated effect of the number of CIs on the number of overseas students, α_1, refers to a population, rather than a country, average effect from 2002 to 2014.

The identification of our empirical model requires that the number of CIs a country has is exogenous. However, the decision to open a CI may depend on various factors. Lien and Oh (2014) argue that GDP, population size, geographical distance to China, and English as a major spoken language are the most important determinants for the establishment of a CI. Furthermore, trade and/or FDI and developing country status are shown to have positive, respectively negative, effects on CI openings. Geographic distance, the usage of English in a country, and developing country status are almost constant over time for countries in our observation period, and are hence controlled for by the country fixed effects in our empirical model. Moreover, as we standardize our outcome variables and the main independent variable by population, and use GDP and trade measures as covariates in our regression models, we effectively control for the most important factors suggested by Lien and Oh (2014) that determine the opening of CIs. Importantly, these factors are also the ones influencing the international mobility of students, as suggested by González et al. (2011).

The number of scholarships provided by the Chinese government to international students and the student exchange programs may rise over time. If such positive change in scholarships or exchange programs is correlated with the evolution of CIs, we may overestimate the effect of CIs on overseas students. However, the grant of Chinese government scholarships to overseas students in China is a fair process. One of the most important determinants is the study proposal or study plan of students.[13] Therefore, all overseas students have the chance to apply for these scholarships and the decision of issuing the scholarship is not correlated with the nationality of students. The aggregate trend in the student exchange programs can be controlled for by the year fixed effects. Nevertheless, it is possible that such programs are organized more often between China and some specific countries. In order to overcome this possible endogeneity issue, we add to the empirical model the region-specific trends in

[13]More information about Chinese government scholarships can be found at: https://www.chinesescholarshipcouncil.com.

one robustness check. Detailed discussion is provided in Section 4.2.2.

We consider in our analysis an unbalanced panel data set of countries that have complete information on all variables. Due to a lack of data on overseas students in China before 2002, our observation period starts in 2002, two years before the establishment of the first CI. Therefore, in our observation period (2002 – 2014), all countries have some years recorded when they did not yet host a CI. We exclude Hong Kong, Macao, and Taiwan from our estimation sample, since it is much easier for people from these entities to learn Chinese (Mandarin), and China also considers these entities as an integral part of the People's Republic of China. Our final estimation sample consists of 2,250 country-year observations for 182 countries in the period 2002 to 2014. Table A-1 in the Appendix lists all countries in our estimation sample, and Table A-2 documents the numbers of countries by the number of CIs they had (ranging from 0 to 107) at the end of the observation period (in 2014). 36% of the countries had no institute, and 34% only one. Countries with 10 or more institutes account for 6.6% of all countries. Table 1 below provides summary statistics for our estimation sample.

Table 1: Summary statistics for the estimation sample

	Observation	Mean	Std. Dev.	Min	Max
ST^{total}	2,250	139.0447	433.4705	0	7,603.1885
ST^{long}	2,250	100.9939	344.0978	0	6,868.5059
ST^{short}	2,250	38.0508	101.1259	0	2,198.2632
CI	2,250	0.1369	0.3094	0	36.5118
$lgdp_{-1}$	2,250	7.9482	1.6017	4.6636	11.6521
$lexp_{-1}$	2,250	12.5428	2.6608	0	16.7837
$limp_{-1}$	2,250	13.2718	2.0636	3.0445	17.4221

Notes: The table shows summary statistics for the full estimation sample, which consists of 182 countries. ST and CI measure the population standardized number of overseas students and CIs from 2002 to 2014, respectively. $(\cdot)_{-1}$ indicates a variable measured in 2001 – 2013, one year prior to the outcome measurement. Means and standard deviations are weighted by a country's population in 2002. For a description of all variables and the data sources used to construct them, see main text.

4 Results

We first present and discuss our main results in Section 4.1. Section 4.2 then discusses potential confounders that may influence both the number of overseas students and Confucius Institutes (CIs) and also provides several robustness checks. In Section 4.3, we explore potential effect heterogeneity.

4.1 Main results

We estimate equation (1) for three outcomes (total overseas students, long- and short-term overseas students) using a fixed-effects regression model. Table 2 presents the estimation results.[14] The coefficient on the core independent variable CI is statistically significant at the 1% significance level and has the expected positive sign for all three outcomes.[15] If the number of CIs increases by 1, equivalent to an average increase by 0.1003 institutes per 1 million population aged 15 – 34[16], the total number of overseas students per 1 million population aged 15 – 34 rises by about 26.7 ($= 0.1003 \times 266.279$), which is about 19.2% of its mean value in the observation period. Similarly, an increase in the number of CIs by 1 leads to a rise in the number of long-term overseas students by 18% and a rise in the number of short-term ones by 22.4% relative to their respective means. Since long-term studies require more enthusiasm and dedication in order to learn both the Chinese language and culture than short-term studies, the smaller effect of CIs on long-term students is not surprising. GDP per capita shows a significant positive effect on total and long-term overseas students, which implies that countries with better economic conditions tend to send more people to China for study, especially for longer term. The effects of exports to and imports from China exert a significant negative effect in some specifications, which seems odd at first glance. However, in all specifications, we also controlled for GDP as an indicator for overall economic conditions.

The results shown in Table 2 suggest a positive average marginal effect of opening CIs on overseas students in China from 2002 to 2014. However, we should interpret these results with caution. First, there is no theory (or compelling argument) to support an exclusively delayed effect of CIs on students. Indeed, it is possible that students who acquire at least some knowledge of Chinese immediately go to China to study, especially short-term. According to the definitions of long- and short-term studies, joining a one-year exchange program in Chinese universities is classified as long-term study and the demand for Chinese language skills may not be high, especially when courses at the Chinese university are given partly in English. Therefore, learning just a little Chinese in the home country may prompt students to go immediately to China for study. Nonetheless, a delayed effect may occur if students

[14] We also estimated our main regression model using a sample which excludes the US, Germany, France, Sweden, and Canada, i.e. countries which closed at least one CI. The results are qualitatively similar (not tabulated).

[15] Our results are different from those of Lin et al. (2016), who find a negative effect of CIs on overseas student numbers. We tried to replicate their estimation to gauge the reason for their divergent findings. Detailed discussion is provided in Appendix C. The analysis indicates that the functional form employed by Lin et al. (2016) may be wrong.

[16] We calculate the equivalent increase in the number of CIs by the population weighted average of (1/population aged 15-34)×1,000,000.

TABLE 2: The effect of CIs abroad on the number of overseas students in China

	Total (1)	Long-term (2)	Short-term (3)
CI	266.279***	181.427***	84.852***
	(85.970)	(69.771)	(18.322)
$lgdp_{-1}$	105.088**	96.219**	8.869
	(48.406)	(42.129)	(8.114)
$lexp_{-1}$	−4.948	−2.355	−2.593*
	(6.437)	(5.348)	(1.504)
$limp_{-1}$	−36.947**	−23.009	−13.937***
	(17.486)	(14.075)	(4.188)
Country FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Mean of dep. var.	139.0447	100.9939	38.0508
Observations	2,250	2,250	2,250
# of countries	182	182	182

Notes: This table shows the main results from fixed-effects regressions of the effect of the number of CIs abroad on the number of overseas students in China from 2002 to 2014. The dependent variables are the number of overseas students in total (column (1)), for long-term study (column (2)), and for short-term study (column (3)) in China per 1,000,000 population aged 15 – 34 of the country of origin. The main independent variable is the number of CIs per 1,000,000 population aged 15 – 34 of a CI host country. Control variables considered in all regressions include country and year fixed effects, the logarithm of GDP per capita of students' country of origin ($lgdp_{-1}$), the logarithm of the absolute value of exports from the students' country of origin to China ($lexp_{-1}$), and the logarithm of the absolute value of imports from China to students' country of origin ($limp_{-1}$). All control variables are measured one year prior to the measurement of outcome variables. Regressions are weighted by the 2002 population of students' country of origin. Standard errors are clustered at country level. *, **, *** denote statistical significance at the 10%, 5% and 1% level.

have mastered a high level of Chinese through long-time study at a CI and decide to go to China for study afterwards. In this case, an additional CI may start to influence the number of students only one or several years after the opening. Second, no information on the size (capacity) of CIs is available, such as the number of courses or teachers. The marginal effect of opening a CI could also rise over time if more learning opportunities are provided by an institute as time progresses. To gauge the relevance of these two arguments, we regress ST_{it} on $CI_{i,t-1}$, and $CI_{i,t-2}$ separately for all three outcome measures. The results are shown in Table A-3 in the Appendix. As it turns out, the effects of CIs on student numbers are still positive, but the size of coefficients becomes larger when we predate the measure of CIs further back in time before the measurement of student numbers. Therefore, there seems to be some delayed effect of opening CIs on the number of overseas students in China, possibly because it requires some time until the learning opportunities at an institute reaches full capacity. This finding also corroborates the analysis in Section 4.2.1 on the pattern of overseas student numbers before and after the opening of a CI.

In our model, we standardize both the number of students and institutes by annual population. It is therefore possible that the number of institutes (or students) does not change, but that the population share of institutes (or students) does change when the

population of a country varies over time. Although changes in population should be minor, we checked whether our results are sensitive to population changes. In order to do this, we standardized the number of institutes and students by the 2002 population aged 15 – 34 of a country, so that changes in both ratios only originate from variations in the number of institutes and students. The results are shown in columns (1) to (3) of Table A-4 in the Appendix. Compared to our main results, the effect of CI on ST is still positive and statistically significant, and the size of coefficient is not materially different. Columns (4) to (6) of Table A-4 show the results if we estimate our model without population weights.[17] The country-average effect turns out to be much smaller, presumably because countries with small populations exert a smaller effect than those with larger populations.[18] In the following analyses, we use the specifications in Table 2 as the baseline and estimate population-average effects of the number of CIs abroad on the number of overseas students in China by weighting the regressions by country populations.

4.2 Robustness checks

The establishment of CIs is unlikely to be random across countries and time. CI openings may, for example, correlate with the demand for Chinese language instruction in host countries, and the willingness and capability of the potential host organization to establish an institute (Lien and Oh, 2014). In our main analysis, we use a fixed effects model to control for country-specific time-invariant observable and unobservable factors. However, country and time fixed effects will not suffice for identification in the presence of spatial variation in unobservables that confound the relationship between CIs and overseas student numbers across countries and time. Below, we therefore address such causality issues in several ways. First, in Section 4.2.1, we check whether the number of overseas students changes significantly before the opening of CIs. Second, in Section 4.2.2, we examine whether our main findings still apply if we control for region-specific trends. Finally, in Section 4.2.3, we check the robustness of our main results to various changes in the estimation sample.

[17]The effect on short-term students in column (6) is statistically insignificant because of an outlier, the Seychelles. The Seychelles opened the first CI only in 2014, and has the smallest population in our sample. If we drop the Seychelles from our estimation sample, and re-estimate the model, the results still show a statistically significant positive effect of the number of CIs on the number of short-term students.

[18]We checked whether countries with large populations show a stronger effect of the number of CIs on the number of students. To do so, we classified countries by the mean (or median) 2002 population into two groups and generated a dummy variable that is 1 if a country's population is above the mean (or median) value, and 0 otherwise. We included the interaction term between this dummy variable and CI in the regression model and estimated a fixed effects model without population weights. The results indicate that the estimated effect of CI on ST tends to be larger for countries with large populations (not tabulated).

4.2.1 Number of overseas students before the opening of a CI

As discussed in Section 3.2, the most important determinants of CI establishment suggested by Lien and Oh (2014) are already considered in our baseline fixed effects model. However, it is still possible that our findings suffer from bias, because of anticipation effects (of closer foreign-country-China ties) or unobserved pre-trends in the demand for overseas study in China that correlate with (or influence outright) the opening of a CI.

To gauge the importance of such a threat to identification, we check whether the number of overseas students changes systematically before the establishment of CIs. For this purpose, we construct a series of dummy variables, $Open_{i,t-j}$, which indicate whether the current year t, when overseas students are observed for a specific country i, coincides with the year in which a CI was opened ($j = 0$), is the jth year before (if $j < 0$), or the jth year after (if $j > 0$) the opening of a CI. We then estimate the following model:

$$ST_{it} = \alpha + \sum_{j=-1}^{n} \beta_j \times Open_{i,t-j} + \boldsymbol{\gamma X_{i,t-1}} + \boldsymbol{d_i} + \boldsymbol{d_t} + \nu_{it}, \tag{2}$$

where ST_{it} is the standardized number of overseas students (total, long-term and short-term) from country i to China in year t. $\sum_{j=-1}^{n} Open_{i,t-j}$, our main explanatory variables are a set of indicator variables capturing whether the current period (calendar year) t is leading an opening of a CI by one year ($j = -1$), coinciding with the period of an opening ($j = 0$), or lagging a CI establishment by one to n years ($j = 1, \ldots, n$). In the empirical analysis, we choose n to be equal to 5 (denoted by $5plus$), meaning that $Open_{t-5plus}$ indicates all periods that are at least 5 years after a CI opening. The reference group (period) consists of years that are at least two years before the first opening of a CI. Since a country can have several institutes that opened in different years, for country i in a certain year t, two or more indicators could be equal to 1.[19] The coefficients β_j measure the relative change in the number of overseas students one year before, in the year of, or several years after the opening of a CI. $\boldsymbol{X_{i,t-1}}$ is a set of control variables that are the same as in our main model in Section 4.1. Vector $\boldsymbol{d_i}$ controls for country fixed effects, and vector $\boldsymbol{d_t}$ (a set of year dummies) controls for country-invariant changes in overseas students between different calendar years. Finally, ν_{it} is an error term.

We are interested in the β_j coefficients. The coefficient β_{-1} on our indicator for the year before the new establishment of a CI measures two important things. First, it captures any anticipatory effect on overseas students. As discussed above, if people anticipate that there

[19] It is also possible that several institutes opened in one year in a particular country. However, we do not differentiate between multiple openings in one year in our model. One year is taken as an opening year if at least one institute is launched in that year.

will be a CI established in their region, they may change their behavior in anticipation of closer cooperation between their country and China and therefore go in larger numbers to China for study. Secondly, and more importantly, it indicates whether there is a systematic change in overseas students one year before a CI opening, which would indicate that CI openings are endogenous to past recent developments in overseas student numbers. β_{-1} should be statistically indifferent from zero if the opening of CIs changes the behavior of people only after the openings occurred, and such openings are not driven by past developments of overseas student numbers.

Estimation results for equation (2) are shown in Table 3. The estimated coefficient on the lead variable $Open_{t+1}$ is statistically insignificant in all three specifications, which suggests no anticipation effect and no endogeneity of CI openings to past recent changes in overseas student numbers. Furthermore, in column (3), the immediate response of overseas students for short-term study in China to a CI opening (the coefficient on $Open_{t-1}$) is statistically significantly positive and of larger magnitude than the coefficient on the lead variable $Open_{t+1}$. However, such an immediate effect is not observable for long-term overseas students (column (2)) and the summary total of overseas students (column (1)). The coefficient on $Open_{t-5plus}$ in columns (1) – (3) is significantly positive, and its magnitude is also the largest, which corroborates our findings in Section 4.1 that there may be a delayed effect of opening CIs on the number of overseas students in China.

We also re-did the analysis, dropping countries from the estimation sample that have no CIs during the whole observation period, so as to have a more homogeneous set of countries in terms of Chinese language learning opportunities (i.e. countries that are ever treated). The results, shown in columns (1) – (3) of Table A-5 in the Appendix, are similar. We also checked whether the decision to open a CI (which may predate the actual opening by more than one year) is influenced by systematic changes in overseas student numbers one to two years prior to the establishment of a CI. In order to do this, we included one more dummy variable, $Open_{t+2}$, to check whether the number of students two years before the opening of a CI changes significantly from earlier years. The estimation results are shown in columns (4) – (6) of Table A-5. The coefficients on both $Open_{t+2}$ and $Open_{t+1}$ are statistically insignificant. Therefore, the decision to establish a CI does not appear to be associated with prior variation in the number of overseas students.

4.2.2 Region-specific trends

The opening of CIs across countries may correlate with region-specific trends in overseas student numbers in China, if the expansion of trade or cultural ties (such as student exchange programs) with China evolved differently across time in different parts of the world. To

TABLE 3: Number of overseas students in China before and after the opening of a CI abroad

	Total (1)	Long-term (2)	Short-term (3)
$Open_{t+1}$	1.189	−1.899	3.088
	(10.724)	(8.129)	(3.077)
$Open_t$	10.426	2.234	8.192
	(14.789)	(10.205)	(4.970)
$Open_{t-1}$	17.243	9.560	7.683*
	(17.606)	(13.660)	(4.322)
$Open_{t-2}$	5.146	1.072	4.073
	(14.753)	(11.935)	(3.980)
$Open_{t-3}$	24.781	15.290	9.491**
	(15.202)	(11.679)	(4.678)
$Open_{t-4}$	23.873	13.180	10.693**
	(15.003)	(11.660)	(4.251)
$Open_{t-5plus}$	75.531**	52.616**	22.914***
	(30.316)	(25.176)	(6.477)
Other controls	Yes	Yes	Yes
Country FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Observations	2,250	2,250	2,250
# of countries	182	182	182

Notes: We analyze the number of overseas students before and after an opening of a CI. The dependent variables are the number of overseas students in total (column (1)), for long-term study (column (2)), and for short-term study (column (3)) in China per 1,000,000 population aged 15 – 34 of the country of origin. The main independent variables are a set of dummies, $Open_{t-j}$. $Open_t$ takes value 1 if a new CI is established in the current year t. The dummy variable $Open_{t+1}$ brings forward an opening of a CI by one year, and dummy variables $Open_{t-1}$ to $Open_{t-5plus}$ capture the response over time of overseas students to the opening of a new CI. Control variables considered in all regressions include country and year fixed effects, the logarithm of GDP per capita of student country of origin ($lgdp_{-1}$), the logarithm of the absolute value of exports from students' country of origin to China ($lexp_{-1}$), and the logarithm of the absolute value of imports from China to students' country of origin ($limp_{-1}$). All control variables are measured one year prior to the measurement of outcome variables. Regressions are weighted by the 2002 population of students' country of origin. Standard errors are clustered at country level. *, **, *** denote statistical significance at the 10%, 5% and 1% level.

gauge the importance of such potentially confounding influences on our baseline results, we re-estimated our specifications for overseas student numbers in China (total, long- and short-term study), now also controlling for linear, or linear-quadratic, region-specific trends.

We first classify countries into 22 regions according to their geographic location.[20] The estimation results are shown in Panel (A) of Table 4. Specifications in columns (1), (3) and (5) control for linear region trends, those in columns (2), (4) and (6) for linear-quadratic region trends. As is evident, the number of CIs abroad still exerts a statistically significant positive effect on the number of overseas students in China. Considering specific trends by regions, classified by geographic locations alone, however, may ignore important economic differences between countries in such regions. Based on their geographic locations and income

[20]These regions are Central America, Northern America, South America, Central Asia, Eastern Asia, South-Eastern Asia, Southern Asia, Western Asia, Eastern Africa, Middle Africa, Northern Africa, Southern Africa, Western Africa, Eastern Europe, Northern Europe, Southern Europe, Western Europe, Australia/New Zealand, Polynesia, Melanesia, Micronesia, and the Caribbean.

TABLE 4: The effect of CIs abroad on the number of overseas students in China when controlling for region-specific trends (linear and non-linear)

	Total		Long-term		Short-term	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel (A): 22 regions						
CI	269.765**	264.660***	209.540**	204.938***	60.225**	59.721**
	(106.664)	(101.215)	(83.745)	(78.946)	(24.742)	(24.313)
$lgdp_{-1}$	73.257	91.408	71.846	87.562	1.411	3.846
	(60.116)	(66.808)	(48.535)	(53.998)	(14.464)	(15.805)
$lexp_{-1}$	−0.500	−2.003	0.052	−1.135	−0.552	−0.869
	(5.310)	(5.313)	(4.450)	(4.428)	(1.227)	(1.256)
$limp_{-1}$	2.954	13.315	−0.025	9.663	2.979	3.652
	(20.035)	(25.680)	(17.227)	(22.162)	(3.970)	(4.852)
Country FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Region-specific trends (linear)	Yes	No	Yes	No	Yes	No
Region-specific trends (linear-quadratic)	No	Yes	No	Yes	No	Yes
Observations	2,250	2,250	2,250	2,250	2,250	2,250
# of countries	182	182	182	182	182	182
Panel (B): 51 regions						
CI	208.699*	208.306**	161.007*	159.316**	47.692**	48.989**
	(107.352)	(103.468)	(83.999)	(80.178)	(23.831)	(23.702)
$lgdp_{-1}$	39.693	59.743	25.984	41.632	13.709	18.111
	(47.275)	(54.506)	(34.594)	(40.190)	(13.172)	(14.841)
$lexp_{-1}$	0.986	1.058	−0.251	−0.067	1.237	1.125
	(4.828)	(5.178)	(3.948)	(4.213)	(1.084)	(1.133)
$limp_{-1}$	13.487	26.312	8.894	19.912	4.593	6.400
	(15.850)	(24.359)	(13.641)	(20.687)	(2.884)	(4.330)
Country FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Region-specific trends (linear)	Yes	No	Yes	No	Yes	No
Region-specific trends (linear-quadratic)	No	Yes	No	Yes	No	Yes
Observations	2,250	2,250	2,250	2,250	2,250	2,250
# of countries	182	182	182	182	182	182

Notes: Specifications in Panel (A) control for linear, respectively linear-quadratic, trends of 22 different country regions of the world, and specifications in Panel (B) consider a larger set of regions (51), constructed based both on the geographic location of countries and their levels of income. Columns (1), (3), and (5) control for region-specific linear time trends, and columns (2), (4), and (6) for region-specific non-linear time trends. The dependent variables are the number of overseas students in total (columns (1) and (2)), for long-term study (columns (3) and (4)), and for short-term study (columns (5) and (6)) in China per 1,000,000 population aged 15 – 34 of the country of origin. The main independent variable is the number of CIs per 1,000,000 population aged 15 – 34 of a CI-host country. Control variables considered in all regressions include country and year fixed effects, the logarithm of GDP per capita of students' country of origin ($lgdp_{-1}$), the logarithm of the absolute value of exports from students' country of origin to China ($lexp_{-1}$), and the logarithm of the absolute value of imports from China to students' country of origin ($limp_{-1}$). All control variables are measured one year prior to the measurement of outcome variables. Regressions are weighted by the 2002 population of students' country of origin. Standard errors are clustered at country level. *, **, *** denote statistical significance at the 10%, 5% and 1% level.

levels, we therefore construct a second regional classification, which consists of 51 groups of high-, upper-middle-, lower-middle- and low-income countries in different parts of the world.[21] The estimation results are shown in Panel (B) of Table 4. Again, we find a significant positive effect of the number of CIs on the number of overseas students.

[21] We employ the World Bank's benchmark of country income levels in 2002. A detailed description of this classification is provided in Section 4.3.2.

4.2.3 Changes in the estimation sample

In this section, we check the robustness of our findings to various changes in the estimation sample that provide for a more homogeneous set of countries to gauge the representativeness of our results and the role of potentially influential observations. First, we drop countries that have the most CIs (top 5%) during the observation period. Second, we drop countries with no CIs during the observation period. Last but not least, we drop countries with no overseas students in China during the observation period. We employ the empirical model of Section 3.2 (baseline specification) for our robustness checks. Table 5 shows the estimation results for the three restricted samples.

First, we drop countries that have the most CIs. At the end of 2014, 10 countries (top 5% of countries) had at least 12 institutes (see Table A-2 in the Appendix). These include Australia, Canada, Germany, France, the UK, Japan, South Korea, Russia, Thailand, and the US. Excluding these 10 countries, we still find a statistically significant sizeable positive effect of the number of CIs abroad on the number of overseas students in China (see columns (1) – (3) of Table 5).

Second, we drop countries that have no institutes during the whole observation period. In our original estimation sample, 66 countries never opened a CI (see Table A-2), but would have some overseas students in China. As shown in columns (4) – (6) of Table 5, however, the means of the number of overseas students in countries with CIs in the observation period is larger than in our full sample. Countries with CIs, on average, hence have more overseas students than countries with no CIs. The size and statistical significance of the estimated effects of CIs on our three student outcome measures for this restricted sample, however, turn out similar to those of our main results in Table 2.

Finally, we exclude countries that have no overseas students in China at all during the estimation observation period. Estimation results are shown in columns (7) – (9) of Table 5. By construction, the mean values of the dependent variables (the number of overseas students) in this restricted sample must exceed those in the full sample. 19 countries in the original full sample have no overseas students. Dropping these countries, we still find a statistically significant positive effect of the number of CIs on the number of overseas students in China (in total and for both long- and short-term study).

4.3 Effect heterogeneity

In this section, we investigate potential effect heterogeneity. For this purpose, we modify equation (1) by adding interaction terms between CI and indicator variables for certain country features. We first explore potential effect heterogeneity between Asian and non-Asian

TABLE 5: The effect of CIs abroad on the number of overseas students in China using various estimation samples

	Dropping countries with most CIs (5%)			Dropping countries with no CIs			Dropping countries with no students		
	Total (1)	Long-term (2)	Short-term (3)	Total (4)	Long-term (5)	Short-term (6)	Total (7)	Long-term (8)	Short-term (9)
CI	188.998***	141.355**	47.643***	267.530***	184.355**	83.175***	264.817***	180.438**	84.379***
	(63.594)	(55.099)	(11.717)	(87.102)	(70.819)	(18.583)	(86.402)	(70.134)	(18.422)
Other controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean of dep. var.	61.1868	49.5570	11.6298	148.2792	107.0227	41.2565	141.1674	102.5357	38.6316
Observations	2,120	2,120	2,120	1,489	1,489	1,489	2,033	2,033	2,033
# of countries	172	172	172	116	116	116	163	163	163

Notes: This table reports regression results for three different estimation samples. The first, used in columns (1) – (3), excludes the top 5% of countries with the most CIs at the end of 2014. The second, used in columns (4) – (6), excludes countries with no CIs during the observation period (2002 – 2014). The third, used in columns (7) – (9), omits countries with no overseas students in China in the observation period. The dependent variables are the number of overseas students in total (columns (1), (4), and (7)), for long-term study (columns (2), (5), and (8)), and for short-term study (columns (3), (6), and (9)) in China per 1,000,000 population aged 15 – 34 of the country of origin. The main independent variable is the number of CIs per 1,000,000 population aged 15 – 34 of a CI-host country. Control variables considered in all regressions include country and year fixed effects, the logarithm of GDP per capita of students' country of origin ($lgdp_{-1}$), the logarithm of the absolute value of exports from students' country of origin to China ($lexp_{-1}$), and the logarithm of the absolute value of imports from China to students' country of origin ($limp_{-1}$). All control variables are measured one year prior to the measurement of outcome variables. Regressions are weighted by the 2002 population of students' country of origin. Standard errors are clustered at country level. *, **, *** denote statistical significance at the 10%, 5% and 1% level.

countries. For Asian countries, we also check for differential effects related to the linguistic distance between Chinese and the major language of CI-host countries. We then consider effect heterogeneity by country income level. Finally, we investigate whether countries with a large population of Chinese migrants are also affected differently by the opening of a CI than countries with few Chinese residents.

4.3.1 Asian and non-Asian countries

People who are geographically close to each other may share more similar customs, values, and cultural features than people who live far away. Geographical closeness may promote international communication and exchange, and hence increase the demand for learning Chinese language and culture in Asian countries more than in distant countries located on other continents. As the cultures are relatively similar, it is easier for other Asian students to live and study in China. Geographical proximity implies that the travel costs from home countries are on average lower. Opening a CI in an Asian country may therefore have a larger effect on the number of overseas students studying in China. We test for such effect heterogeneity by interacting CI with the indicator variable $Asian$, which equals 1 if a country is an Asian country and 0 otherwise. There are 45 Asian and 137 non-Asian countries in our sample.

Columns (1), (3), and (5) of Table 6 show the regression results for our three student outcome measures. As is evident, CIs exert a statistically significant positive effect on all three overseas student counts, both in Asian and in non-Asian countries, but the effect is much larger in magnitude in the former than in the latter.

Language is generally regarded as the carrier of human culture. If the linguistic distance, i.e. dissimilarity between languages, is smaller between countries, it is easier for their populations to learn the other's language and culture. Previous research has produced evidence that linguistic distance has a negative effect on international migration flows (Adserà and Pytliková, 2015; Belot and Hatton, 2012) as well as bilateral trade volumes (Hutchinson, 2005; Isphording and Otten, 2013). If a country's language has the same roots as Chinese, it is much easier for people from that country to learn Chinese, and their probability of going to China for study is therefore likely to be higher. The effect of CIs on the number of overseas students from countries with close linguistic distance to Chinese may hence well be larger than for other countries. Chinese is a member of the Sino-Tibetan language family. One of the official languages of Singapore is Chinese, and many Singapore residents are ethnic Chinese. The official languages of Bhutan and Myanmar also belong to the Sino-Tibetan language family. We generate a dummy variable for these three countries, $Sino\text{-}Tibetan$, and add, in addition to the interaction between CI and $Asian$, a triple interaction of CI,

TABLE 6: The effect of CIs abroad on the number of overseas students in China from Asian and non-Asian countries

	Total		Long-term		Short-term	
	(1)	(2)	(3)	(4)	(5)	(6)
CI	165.300***	164.950***	100.680***	100.641***	64.620***	64.309***
	(35.700)	(35.501)	(28.167)	(28.151)	(11.007)	(10.881)
CI × *Asian*	911.073***	896.363***	728.532***	726.905***	182.540***	169.458***
	(196.827)	(202.091)	(162.397)	(164.956)	(45.449)	(48.109)
CI × *Asian* × *Sino-Tibetan*		1087.626***		120.343		967.282***
		(206.052)		(168.243)		(48.244)
Other controls	Yes	Yes	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,250	2,250	2,250	2,250	2,250	2,250
# of countries	182	182	182	182	182	182

Notes: Columns (1), (3), and (5) report estimation results of the heterogeneous effect of CIs abroad on the number of overseas students in China in Asian and non-Asian countries. Columns (2), (4), and (6) report estimates of potential effect heterogeneity among Asian countries by their linguistic distance to Chinese. The dependent variables are the number of overseas students in total (columns (1) and (2)), for a long-term study stay (columns (3) and (4)), and for a short-term study stay (columns (5) and (6)) in China per 1,000,000 population aged 15 – 34 of the country of origin. The main independent variable is the number of CIs per 1,000,000 population aged 15 – 34 of a CI-host country. *Asian* is a dummy variable for Asian countries, and *Sino-Tibetan* a dummy variable that takes value 1 if one of a country's official languages belongs to the Sino-Tibetan language family, and 0 otherwise. Control variables considered in all regressions include country and year fixed effects, the logarithm of GDP per capita of students' country of origin ($lgdp_{-1}$), the logarithm of the absolute value of exports from students' country of origin to China ($lexp_{-1}$), and the logarithm of the absolute value of imports from China to students' country of origin ($limp_{-1}$). All control variables are measured one year prior to the measurement of outcome variables. Regressions are weighted by the 2002 population of students' country of origin. Standard errors are clustered at country level. *, **, *** denote statistical significance at the 10%, 5% and 1% level.

Asian, and *Sino-Tibetan* to our set of controls. The coefficient on this triple interaction term captures the differential effect of CIs by the linguistic closeness of CI-host countries in Asia. The estimation results are shown in columns (2), (4), and (6) of Table 6. For the total number of overseas students, and for those on short-term study, the estimated coefficient on the triple interaction term is statistically significant and positive, and larger than the estimated coefficient on the double interaction term, which suggests that the linguistic distance to Chinese plays a crucial role for the impact of CIs on the number of overseas students in China, at least for short-term study.[22]

4.3.2 High- and low-income countries

In this section, we investigate whether the effect of CIs abroad on the number of overseas students in China differs between countries with different income levels. Lin et al. (2016)

[22]The variation in *Sino-Tibetan* is small, because only three Asian countries have an official language that is considered Sino-Tibetan. We generated another dummy variable that is 1 if at least one minority group of residents in a country speak a Sino-Tibetan language, and 0 otherwise. Countries meeting this requirement include Singapore, Myanmar, Bhutan, Nepal, India, Bangladesh, Thailand, Lao People's Democratic Republic, Cambodia, and Vietnam. With the exception of Singapore, Myanmar and Bhutan, the official language of these countries is not Sino-Tibetan. Only ethnic residents speak a Sino-Tibetan language. Estimation results (not tabulated) are similar to those reported in Table 6. For short-term students, the size of the coefficient on the triple interaction is much smaller than the corresponding coefficient in column (6) of Table 6, since the average linguistic distance to Chinese in these 10 countries is larger than the average distance between Chinese and the languages in Singapore, Myanmar, and Bhutan.

argue that opening a CI in developing countries can motivate more students from these countries to study in China than in developed countries. We use the World Bank's benchmark in 2002, which classified countries into high-, upper-middle-, lower-middle-, and low-income countries (H-, UM-, LM-, and L-income countries, respectively).[23] 39 countries in our estimation sample are classified as H-income countries, 30 as UM-income, 51 as LM-income and 62 as L-income. We generate a dummy variable for relatively high-income countries, *high income*, which is equal to 1 if the country is classified as a H- or UM-income country and 0 otherwise, and include the interaction between this dummy variable and CI in our regression model. Table 7 shows the results.

TABLE 7: The effect of CIs abroad on the number of overseas students in China in high- and low-income countries

	Total	Long-term	Short-term
	(1)	(2)	(3)
CI	516.748***	420.921***	95.827**
	(167.725)	(142.477)	(38.976)
$CI \times$ *high income*	−294.709	−281.796*	−12.913
	(185.635)	(152.695)	(45.698)
Other controls	Yes	Yes	Yes
Country FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Observations	2,250	2,250	2,250
# of countries	182	182	182

Notes: The dependent variables are the number of overseas students in total (column (1)), for a long-term study stay (column (2)) and for a short-term study stay (column (3)) in China per 1,000,000 population aged 15 – 34 of the country of origin. The main independent variable is the number of CIs per 1,000,000 population aged 15 – 34 of a CI-host country. *high income* is a dummy variable that equals 1 if the country is classified as H- or UM-income country, and 0 otherwise. Control variables considered in all regressions include country and year fixed effects, the logarithm of GDP per capita of students' country of origin ($lgdp_{-1}$), the logarithm of the absolute value of exports from students' country of origin to China ($lexp_{-1}$), and the logarithm of the absolute value of imports from China to students' country of origin ($limp_{-1}$). All control variables are measured one year prior to the measurement of outcome variables. Regressions are weighted by the 2002 population of students' country of origin. Standard errors are clustered at country level. *, **, *** denote statistical significance at the 10%, 5% and 1% level.

For all three specifications, the estimated coefficient on CI remains statistically significant and positive. Furthermore, the estimated coefficient on the interaction term is negatively signed, indicating that the opening of a CI in countries with a relatively high income level tends to exert a smaller effect on the number of overseas students than an opening in poorer countries. However, this differential effect is statistically significant only for long-term stu-

[23] A detailed classification can be found on World Bank's web-page at: https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups. Three countries (Montenegro, Serbia, and South Sudan) were not classified by the World Bank to any income group in 2002. Therefore, for Montenegro and Serbia we use their first classification in 2006 and for South Sudan the classification in 2011. During our observation period, China was classified as a lower-middle-income country before 2010 and thereafter as an upper-middle-income country.

dents (see column (2) of Table 7). China has been classified as a UM- income country in 2010 and continues to grow very rapidly economically. For students from L- and LM-income countries, therefore, China appears to be a favorable destination for overseas study, offering better economic prospects than their country of origin. It is less expensive to study in China than in H-income countries. Students from UM- and H-income countries, by contrast, may have more destination choices for their overseas study. They may also have more opportunities provided by official exchange programs or private language institutions for learning Chinese or studying Chinese culture. What is more, tuition fees in China may not be an obstacle for them. For all these reasons, opening a CI in a relatively high-income country may exert a smaller effect on oversea student numbers in China.

4.3.3 Countries with large or small population shares of Chinese

Countries with a large number of Chinese immigrants may have closer relationships and more extensive ties with China, including education exchanges. People from such countries may also have more chances to interact with Chinese people, learn the Chinese language, and get to know the country's culture. Therefore, as the share of Chinese immigrants becomes larger, the demand for learning the Chinese language and culture could increase, and the outflow of natives to China may also rise. Opening CIs in countries with many Chinese immigrants can lead to large increases in overseas students in China from such countries, if there is pent-up demand for learning Chinese and for overseas study in China. If there is no such pent-up demand, however, because existing opportunities for learning Chinese in these countries are abundant, as are education exchange programs, the effects of opening a CI on overseas student numbers from such countries may also fall short of the effects CI openings have in countries with fewer Chinese immigrants. In this section, we test for differential effects of CIs on overseas student numbers between countries hosting many, respectively few, Chinese immigrants.

Data on Chinese immigrants are available only for a smaller group of countries and for selective years. We use stock data on Chinese immigrants in 119 countries in the year 2000. For these countries, we calculate the population share of Chinese immigrants, defined as the number of Chinese immigrants per 1,000,000 population aged 15 – 34 in 2000, and generate the dummy variable, *high Chinese migration*, which is equal to 1 if a country's population share of Chinese immigrants is above the median of the 119 countries, and 0 otherwise. Re-estimating equation (1) using this sub-sample of countries for total, long- and short-term overseas students (see columns (1), (3), and (5) of Table 8) still produces a sizeable positive effect of CIs on the number of overseas students. Including the interaction between CI and *high Chinese migration* in the regression model still delivers a significant positive

TABLE 8: The effect of CIs abroad on the number of overseas students in China among countries with large, respectively small, Chinese immigrant populations

	Total		Long-term		Short-term	
	(1)	(2)	(3)	(4)	(5)	(6)
CI	279.282***	320.468**	191.312**	240.534**	87.970***	79.934*
	(93.059)	(134.268)	(75.738)	(100.770)	(19.796)	(40.575)
CI × *high Chinese migration*		−47.915		−57.264		9.349
		(160.802)		(121.905)		(45.890)
Other controls	Yes	Yes	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,509	1,509	1,509	1,509	1,509	1,509
# of countries	119	119	119	119	119	119

Notes: Columns (1), (3), and (5) show results from re-estimating equation (1) in Section 3.2 (baseline specification) for the restricted sample of 119 countries with information on Chinese immigrant populations in the year 2000. Columns (2), (4), and (6) show results of specifications with interaction terms. The dependent variables are the number of overseas students in total (columns (1) and (2)), for a long-term study stay (columns (3) and (4)), and for a short-term study stay (columns (5) and (6)) in China per 1,000,000 population aged 15 – 34 of the country of origin. The main independent variable is the number of CIs per 1,000,000 population aged 15 – 34 of a CI-host country. The dummy variable *high Chinese migration* takes value 1 if a country has an above the median population share of Chinese immigrants in 2000 (per 1,000,000 population aged 15 – 34), and 0 otherwise. Control variables considered in all regressions include country and year fixed effects, the logarithm of GDP per capita of students' country of origin ($lgdp_{-1}$), the logarithm of the absolute value of exports from students' country of origin to China ($lexp_{-1}$), and the logarithm of the absolute value of imports from China to students' country of origin ($limp_{-1}$). All control variables are measured one year prior to the measurement of outcome variables. Regressions are weighted by the 2002 population of students' country of origin. Standard errors are clustered at country level. *, **, *** denote statistical significance at the 10%, 5% and 1% level.

coefficient estimate of CI. Estimated coefficients on the interaction term vary in sign, but lack significance in all regressions (see columns (2), (4), and (6) of Table 8).[24] The effects of CIs on the number of overseas students, therefore, do not seem to differ systematically between countries with large, and those with small Chinese immigrant populations.

5 Conclusion

The number of overseas students in China has risen substantially over the last few decades, and so did the number of CIs abroad, since Hanban's establishment of the first CI in 2004. Empirical research on the effects of CIs abroad is still scarce but growing. Little research, however, has been conducted on the impact of CIs abroad on overseas students in China. Studies exploring the causal effects of language institutes on different outcomes also remain rare. Using novel data for 182 countries in the period 2002 to 2014, this study is the first to provide a comprehensive analysis of the effects of CIs abroad on the number of overseas students in China.

Using official and self-compiled data on CIs and overseas students in China by countries of origin, we find evidence in fixed-effects regressions for a positive effect of CIs abroad on the number of overseas students in China. Aiding China's soft power, such increases in overseas

[24] As an alternative, we also employed the population share of Chinese immigrants, a continuous measure for Chinese immigrant intensity of countries, to construct the interaction term. The results are similar (not tabulated).

student numbers may also aid bilateral trade between CI-host countries and China, and thus provide one potential explanation of the positive association between CIs abroad and CI-host countries' trade intensity with China, as found in the literature. Our findings proved robust in several sensitivity checks. Effect heterogeneity analyses furthermore showed that the expansionary effect of CIs on overseas student numbers is much more pronounced for countries in Asia, especially those with an official language close to Chinese, and for countries with low or lower-middle income levels. Overall, therefore, our findings suggest that language institutes abroad can strongly promote bilateral student exchange. The economic benefits of such exchange may be both large and lasting, and provide a fruitful area for future research.

References

Adserà, A. and Pytliková, M. (2015). The role of language in shaping international migration. *The Economic Journal*, 125(586):F49–F81.

Akhtaruzzaman, M., Berg, N., and Lien, D. (2017). Confucius Institutes and FDI flows from China to Africa. *China Economic Review*, 44:241–252.

Belot, M. V. and Hatton, T. J. (2012). Immigrant selection in the OECD. *Scandinavian Journal of Economics*, 114(4):1105–1128.

Berman, E., Lang, K., and Siniver, E. (2003). Language-skill complementarity: returns to immigrant language acquisition. *Labour Economics*, 10(3):265–290.

Bessey, D. (2012). International student migration to Germany. *Empirical Economics*, 42(1):345–361.

Blundell, R. and Bond, S. (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics*, 87(1):115–143.

Chen, Q. (2014). *Gaoji jiliang jingjixue ji Stata yingyong [Advanced econometrics and application with Stata]*. Higher Education Press, Beijing, 2nd edition.

Chiswick, B. R. and Miller, P. W. (1995). The endogeneity between language and earnings: International analyses. *Journal of Labor Economics*, 13(2):246–288.

Čok, L. and Novak-Lukanovič, S. (2005). Language as social cohesion and human capital. In Vodovnik, Z., editor, *Intellectual capital and knowledge management: Proceedings of the 5th International Conference of the Faculty of Management Koper, University of Primorska, 18–20 November 2004, Congress Centre Bernardin, Portorož, Slovenia*, pages 79–89. University of Primorska Faculty of Management, Koper.

Collins, F. L. (2008). Bridges to learning: international student mobilities, education agencies and inter-personal networks. *Global Networks*, 8(4):398–417.

Confucius Institute Headquarters (Hanban) (2018). *Confucius Institute Annual Development Report 2018*. Hanban, Beijing.

Department of Policy Planning of the Ministry of Foreign Affairs of the People's Republic of China (2003). *China's Foreign Affairs*. World Affairs Press, Beijing.

Dustmann, C. and Fabbri, F. (2003). Language proficiency and labour market performance of immigrants in the UK. *The Economic Journal*, 113(489):695–717.

Dustmann, C. and van Soest, A. (2001). Language fluency and earnings: Estimation with misclassified language indicators. *Review of Economics and Statistics*, 83(4):663–674.

Findlay, A. M. (2011). An assessment of supply and demand-side theorizations of international student mobility. *International Migration*, 49(2):162–190.

Ghosh, S., Lien, D., and Yamarik, S. (2017). Does the Confucius Institute network impact cultural distance? A panel data analysis of cross-border flows in and out of China. *Asian Economic Journal*, 31(3):299–323.

González, C. R., Mesanza, R. B., and Mariel, P. (2011). The determinants of international student mobility flows: an empirical study on the Erasmus programme. *Higher education*, 62(4):413–430.

Grenier, G. (2015). The value of language skills. *IZA World of Labor*, 205:1–10.

Groot, G. (2018). Cool Japan versus the China threat: Does Japan's popular culture success mean more soft power? In Hashimoto, K., editor, *Japanese Language and Soft Power in Asia*, pages 15–41. Palgrave Macmillan, Singapore.

Huber, M. and Uebelmesser, S. (2019). Presence of language-learning opportunities and migration. CESifo Working Paper No. 7569.

Hutchinson, W. K. (2005). "Linguistic distance" as a determinant of bilateral trade. *Southern Economic Journal*, 72(1):1–15.

Hwang, S.-S., Xi, J., and Cao, Y. (2010). The conditional relationship between English language proficiency and earnings among US immigrants. *Ethnic and Racial Studies*, 33(9):1620–1647.

Isphording, I. E. and Otten, S. (2013). The costs of Babylon - Linguistic distance in applied economics. *Review of International Economics*, 21(2):354–369.

King, R. and Ruiz-Gelices, E. (2003). International student migration and the European 'year abroad': Effects on European identity and subsequent migration behaviour. *International Journal of Population Geography*, 9(3):229–252.

Kogut, B. and Singh, H. (1988). The effect of national culture on the choice of entry mode. *Journal of International Business Studies*, 19(3):411–432.

Lien, D. and Co, C. Y. (2013). The effect of Confucius Institutes on US exports to China: A state level analysis. *International Review of Economics and Finance*, 27:566–571.

Lien, D., Ghosh, S., and Yamarik, S. (2014). Does the Confucius institute impact international travel to China? A panel data analysis. *Applied Economics*, 46(17):1985–1995.

Lien, D. and Lo, M. (2017). Economic impacts of cultural institutes. *The Quarterly Review of Economics and Finance*, 64:12–21.

Lien, D., Lo, M., and Bojanic, D. (2019). Asymmetric effects of cultural institutes on trade and foreign direct investment. *The World Economy*, 42(5):1520–1553.

Lien, D. and Miao, L. (2018). Effects of Confucius Institutes on China's higher education exports: Evidence from Chinese Partner Universities. *International Review of Economics and Finance*, 57:134–143.

Lien, D. and Oh, C. H. (2014). Determinants of the Confucius Institute establishment. *The Quarterly Review of Economics and Finance*, 54(3):437–441.

Lien, D., Oh, C. H., and Selmier, W. T. (2012). Confucius institute effects on China's trade and FDI: Isn't it delightful when folks afar study Hanyu? *International Review of Economics and Finance*, 21(1):147–155.

Lien, D., Yao, F., and Zhang, F. (2017). Confucius Institute's effects on international travel to China: do cultural difference or institutional quality matter? *Applied Economics*, 49(36):3669–3683.

Lien, D., Yao, F., and Zhang, F. (2018). Confucius Institute's effects on China's higher education exports: A perspective from cultural difference and institutional quality. Department of Economics Working Paper Series No. 18-04, West Virginia University.

Lin, H., Xie, Z., and Zheng, R. (2016). Kongzi xueyuan shifou cujin le haiwai xuesheng laihua liuxue – Jiyu 40 ge guojia 2004 $\sim$ 2014 nian mianban shuju de shizheng jianyan [Does Confucius Institute promote overseas students to study in China – An empirical test based on panel data of 40 countries from 2004 to 2014]. *Guoji Shangwu – Duiwai Jingji Maoyi Daxue Xuebao [International Business]*, 5:52–65.

Lo, J. T. and Pan, S. (2016). Confucius Institutes and China's soft power: practices and paradoxes. *A Journal of Comparative and International Education*, 46(4):512–532.

Miao, L. and Chen, C. (2015). Konzi xueyuan dui woguo gaodeng jiaoyu chukou de yingxiang – Jiyu zhuyao guojia mianban shuju de shizheng yanjiu [Impact of Confucius Institutes on China's education export: An empirical study based on panel data of major countries]. *Guoji Shangwu – Duiwai Jingji Maoyi Daxue Xuebao [International Business]*, 6:27–35.

Nye, J. S. (1990). Soft power. *Foreign Policy*, 80:153–171.

Nye, J. S. (2005). The rise of China's soft power. Wall Street Journal. https://www.wsj.com/articles/SB113580867242333272.

Nye, J. S. (2017). Soft power: the origins and political progress of a concept. *Palgrave Communications*, pages 1–3. DOI: 10.1057/palcomms.2017.8.

Oh, C. H., Selmier, W. T., and Lien, D. (2011). International trade, foreign direct investment, and transaction costs in languages. *The Journal of Socio-Economics*, 40(6):732–735.

Qi, J., Li, L., and Yang, L. (2012). Zhongguo ofdi de quwei xuanze: Jiyu wenhua juli de menkan xiaoying yu jianyan [Location choice of Chinese OFDI: Based on the threshold effect and test of cultural distance]. *Guoji Maoyi Wenti [Journal of International Trade]*, 12:137–147.

Roodman, D. (2009). How to do xtabond2: An introduction to difference and system GMM in Stata. *The stata journal*, 9(1):86–136.

Roy, A., Newman, A., Ellenberger, T., and Pyman, A. (2019). Outcomes of international student mobility programs: a systematic review and agenda for future research. *Studies in Higher Education*, 44(9):1630–1644.

Appendix A Tables Cited in the Main Text

TABLE A-1: Countries in the estimation sample

Country	Code	Asia	Income level	Sino-Tibetan	Country	Code	Asia	Income level	Sino-Tibetan
Aruba	abw	No	H	No	Afghanistan	afg	Yes	L	No
Angola	ago	No	L	No	Albania	alb	No	LM	No
United Arab Emirates	are	Yes	H	No	Argentina	arg	No	UM	No
Armenia	arm	Yes	LM	No	Antigua and Barbuda	atg	No	H	No
Australia	aus	No	H	No	Austria	aut	No	H	No
Azerbaijan	aze	Yes	L	No	Burundi	bdi	No	L	No
Belgium	bel	No	H	No	Benin	ben	No	L	No
Burkina Faso	bfa	No	L	No	Bangladesh	bgd	Yes	L	No
Bulgaria	bgr	No	LM	No	Bahrain	bhr	Yes	H	No
Bahamas	bhs	No	H	No	Bosnia and Herzegovina	bih	No	LM	No
Belarus	blr	No	LM	No	Belize	blz	No	UM	No
Bolivia	bol	No	LM	No	Brazil	bra	No	LM	No
Barbados	brb	No	H	No	Brunei Darussalam	brn	Yes	H	No
Bhutan	btn	Yes	L	Yes	Botswana	bwa	No	UM	No
Central African Republic	caf	No	L	No	Canada	can	No	H	No
Switzerland	che	No	H	No	Chile	chl	No	UM	No
Ivory Coast	civ	No	L	No	Cameroon	cmr	No	L	No
Democratic Republic of the Congo	cod	No	L	No	Congo	cog	No	L	No
Colombia	col	No	LM	No	Comoros	com	No	L	No
Cape Verde	cpv	No	LM	No	Costa Rica	cri	No	UM	No
Cuba	cub	No	LM	No	Cyprus	cyp	No	H	No
Czech Republic	cze	No	UM	No	Germany	deu	No	H	No
Djibouti	dji	No	LM	No	Denmark	dnk	No	H	No
Dominican Republic	dom	No	LM	No	Algeria	dza	No	LM	No
Ecuador	ecu	No	LM	No	Egypt	egy	No	LM	No
Spain	esp	No	H	No	Estonia	est	No	UM	No
Ethiopia	eth	No	L	No	Finland	fin	No	H	No
Fiji	fji	No	LM	No	France	fra	No	H	No
Federated States of Micronesia	fsm	No	LM	No	Gabon	gab	No	UM	No
United Kingdom	gbr	No	H	No	Georgia	geo	Yes	L	No
Ghana	gha	No	L	No	Guinea	gin	No	L	No
Gambia	gmb	No	L	No	Guinea-Bissau	gnb	No	L	No
Equatorial Guinea	gnq	No	L	No	Greece	grc	No	H	No
Grenada	grd	No	UM	No	Guatemala	gtm	No	LM	No
Guyana	guy	No	LM	No	Honduras	hnd	No	LM	No
Croatia	hrv	No	UM	No	Haiti	hti	No	L	No
Hungary	hun	No	UM	No	Indonesia	idn	Yes	L	No
India	ind	Yes	L	No	Ireland	irl	No	H	No
Islamic Republic of Iran	irn	Yes	LM	No	Iraq	irq	Yes	LM	No
Iceland	isl	No	H	No	Israel	isr	Yes	H	No
Italy	ita	No	H	No	Jamaica	jam	No	LM	No
Jordan	jor	Yes	LM	No	Japan	jpn	Yes	H	No
Kazakhstan	kaz	Yes	LM	No	Kenya	ken	No	L	No
Kyrgyzstan	kgz	Yes	L	No	Cambodia	khm	Yes	L	No
Kiribati	kir	No	LM	No	Republic of Korea	kor	Yes	H	No
Kuwait	kwt	Yes	H	No	Lao People's Democratic Republic	lao	Yes	L	No
Lebanon	lbn	Yes	UM	No	Liberia	lbr	No	L	No
Libyan Arab Jamahiriya	lby	No	UM	No	Saint Lucia	lca	No	UM	No
Sri Lanka	lka	Yes	LM	No	Lesotho	lso	No	L	No
Lithuania	ltu	No	UM	No	Luxembourg	lux	No	H	No

Notes: Table is continued on next page.

Table A-1: Countries in the estimation sample (continued)

Country	Code	Asia	Income level	Sino-Tibetan	Country	Code	Asia	Income level	Sino-Tibetan
Latvia	lva	No	UM	No	Morocco	mar	No	LM	No
Republic of Moldova	mda	No	L	No	Madagascar	mdg	No	L	No
Maldives	mdv	Yes	LM	No	Mexico	mex	No	UM	No
Macedonia	mkd	No	LM	No	Mali	mli	No	L	No
Malta	mlt	No	H	No	Myanmar	mmr	Yes	L	Yes
Montenegro	mne	No	UM	No	Mongolia	mng	Yes	L	No
Mozambique	moz	No	L	No	Mauritania	mrt	No	L	No
Mauritius	mus	No	UM	No	Malawi	mwi	No	L	No
Malaysia	mys	Yes	UM	No	Namibia	nam	No	LM	No
Niger	ner	No	L	No	Nigeria	nga	No	L	No
Nicaragua	nic	No	L	No	Netherlands	nld	No	H	No
Norway	nor	No	H	No	Nepal	npl	Yes	L	No
New Zealand	nzl	No	H	No	Oman	omn	Yes	UM	No
Pakistan	pak	Yes	L	No	Panama	pan	No	UM	No
Peru	per	No	LM	No	Philippines	phl	Yes	LM	No
Papua New Guinea	png	No	L	No	Poland	pol	No	UM	No
Puerto Rico	pri	No	H	No	Portugal	prt	No	H	No
Paraguay	pry	No	LM	No	Palestine	pse	Yes	LM	No
Qatar	qat	Yes	H	No	Romania	rou	No	LM	No
Russia Federation	rus	No	LM	No	Rwanda	rwa	No	L	No
Saudi Arabia	sau	Yes	UM	No	Sudan	sdn	No	L	No
Senegal	sen	No	L	No	Singapore	sgp	Yes	H	Yes
Solomon Islands	slb	No	L	No	Sierra Leone	sle	No	L	No
El Salvador	slv	No	LM	No	Somalia	som	No	L	No
Serbia	srb	No	UM	No	South Sudan	ssd	No	LM	No
Sao Tome and Principe	stp	No	L	No	Suriname	sur	No	LM	No
Slovakia	svk	No	UM	No	Slovenia	svn	No	H	No
Sweden	swe	No	H	No	Seychelles	syc	No	UM	No
Syrian Arab Republic	syr	Yes	LM	No	Chad	tcd	No	L	No
Togo	tgo	No	L	No	Thailand	tha	Yes	LM	No
Tajikistan	tjk	Yes	L	No	Turkmenistan	tkm	Yes	LM	No
Timor-Leste	tls	Yes	L	No	Tonga	ton	No	LM	No
Trinidad and Tobago	tto	No	UM	No	Tunisia	tun	No	LM	No
Turkey	tur	Yes	LM	No	United Republic of Tanzania	tza	No	L	No
Uganda	uga	No	L	No	Ukraine	ukr	No	LM	No
Uruguay	ury	No	UM	No	United States	usa	No	H	No
Uzbekistan	uzb	Yes	L	No	Saint Vincent & the Grenadines	vct	No	LM	No
Venezuela	ven	No	UM	No	Vietnam	vnm	Yes	L	No
Vanuatu	vut	No	LM	No	Samoa	wsm	No	LM	No
Yemen	yem	Yes	L	No	South Africa	zaf	No	LM	No
Zambia	zmb	No	L	No	Zimbabwe	zwe	No	L	No

Notes: This table shows the 182 countries that are used in this study, their International Organization for Standardization (ISO) alpha-3 codes, geographic locations (Asia or non-Asia), income level (low (L), lower-middle (LM), upper-middle (UM), and high (H) income), and use of Sino-Tibetan language as an official language (yes/no). These countries are sorted alphabetically according to countries' ISO alpha-3 codes.

TABLE A-2: Number of CIs per country in 2014

# of CIs in 2014 (1)	Frequency (2)	Percent (3)	Cumulative percent (4)
0	66	36.26	36.26
1	62	34.07	70.33
2	22	12.09	82.42
3	9	4.95	87.36
4	6	3.30	90.66
5	3	1.65	92.31
6	2	1.10	93.41
10	1	0.55	93.96
11	1	0.55	94.51
12	2	1.10	95.60
13	2	1.10	96.70
16	2	1.10	97.80
17	1	0.55	98.35
20	1	0.55	98.90
25	1	0.55	99.45
107	1	0.55	100.00
Total	182	100.00	100.00

Notes: This table shows the number of CIs countries in our estimation sample have at the end of our observation period (2014). Column (1) shows the number of CIs and column (2) shows how many countries have the corresponding number of institutes. Columns (3) and (4) present the percent and cumulative percent of countries that have a certain (maximum) number of CIs.

TABLE A-3: The effect of CIs abroad on the number of overseas students in China when using lagged CI variables

	Total			Long-term			Short-term		
	baseline t	$t-1$	$t-2$	baseline t	$t-1$	$t-2$	baseline t	$t-1$	$t-2$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
CI	266.279***			181.427***			84.852***		
	(85.970)			(69.771)			(18.322)		
CI_{-1}		300.935***			202.593***			98.342***	
		(91.337)			(75.014)			(18.691)	
CI_{-2}			314.688***			209.118***			105.571***
			(89.739)			(72.889)			(20.006)
$lgdp_{-1}$	105.088**	111.011**	108.985**	96.219**	99.629**	97.615**	8.869	11.382	11.370
	(48.406)	(48.468)	(48.003)	(42.129)	(42.370)	(41.936)	(8.114)	(7.885)	(8.166)
$lexp_{-1}$	−4.948	−4.750	−4.374	−2.355	−2.267	−2.067	−2.593*	−2.483*	−2.307*
	(6.437)	(6.056)	(6.033)	(5.348)	(5.144)	(5.162)	(1.504)	(1.334)	(1.298)
$limp_{-1}$	−36.947**	−33.183*	−35.795*	−23.009	−20.826	−22.968	−13.937***	−12.357***	−12.827***
	(17.486)	(17.867)	(20.012)	(14.075)	(14.439)	(15.948)	(4.188)	(4.173)	(4.680)
Country FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,250	2,250	2,250	2,250	2,250	2,250	2,250	2,250	2,250
# of countries	182	182	182	182	182	182	182	182	182

Notes: This table shows the effect of the number of CIs on the number of overseas students from 2002 to 2014. The dependent variables are the number of overseas students in total (columns (1) – (3)), for a long-term study stay (columns (4) – (6)), and for a short-term study stay (columns (7) – (9)) in China per 1,000,000 population aged 15 – 34 of the country of origin. The main independent variables are the number of CIs per 1,000,000 population aged 15 – 34 of a CI-host country, measured in period t (columns (1), (4), and (7)), $t-1$ (columns (2), (5), and (8)), and $t-2$ (columns (3), (6), and (9)).Control variables considered in all regressions include country and year fixed effect, the logarithm of GDP per capita of students' country of origin ($lgdp_{-1}$), the logarithm of the absolute value of exports from students' country of origin to China ($lexp_{-1}$), and the logarithm of the absolute value of imports from China to students' country of origin ($limp_{-1}$). All control variables are measured one year prior to the measurement of outcome variables. Regressions are weighted by the 2002 population of students' country of origin. Standard errors are clustered at country level. *, **, *** denote statistical significance at the 10%, 5% and 1% level.

TABLE A-4: The effect of CIs abroad on the number of overseas students in China when standardizing by 2002 populations and using unweighted regressions

	Standardize with population in 2002			Without weighting by population		
	Total (1)	Long-term (2)	Short-term (3)	Total (4)	Long-term (5)	Short-term (6)
CI	242.613***	160.062***	82.552***	29.767**	28.865***	0.903
	(64.889)	(55.615)	(13.224)	(12.142)	(10.559)	(5.875)
$lgdp_{-1}$	114.856**	103.744**	11.112	126.249	112.567	13.682
	(51.559)	(45.542)	(8.406)	(181.472)	(163.072)	(20.682)
$lexp_{-1}$	−2.289	−0.319	−1.970	−23.280**	−17.234*	−6.046***
	(7.564)	(6.420)	(1.626)	(10.585)	(9.415)	(2.281)
$limp_{-1}$	−29.503	−17.268	−12.235**	−2.524	12.247	−14.770*
	(20.516)	(16.576)	(4.973)	(35.492)	(31.534)	(8.198)
Country FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Mean of dep. var.	139.3492	101.7785	37.5708	223.3512	178.892	44.4592
Observations	2,250	2,250	2,250	2,250	2,250	2,250
# of countries	182	182	182	182	182	182

Notes: The dependent variables are the number of overseas students in total (columns (1) and (4)), for a long-term study stay (columns (2) and (5)), and for a short-term study stay (columns (3) and (6)) in China per 1,000,000 population aged 15 – 34 of the country of origin. The main independent variable is the number of CIs per 1,000,000 population aged 15 – 34 of a CI-host country. In columns (1) – (3) we standardize both the dependent and main independent variable by the 2002 population of students' country of origin, and in columns (4) – (6) we standardize these variables by corresponding annual populations. Control variables considered in all regressions include country and year fixed effects, the logarithm of GDP per capita of students' country of origin ($lgdp_{-1}$), the logarithm of the absolute value of exports from students' country of origin to China ($lexp_{-1}$), and the logarithm of the absolute value of imports from China to students' country of origin ($limp_{-1}$). All control variables are measured one year prior to the measurement of outcome variables. Regressions are weighted by the 2002 population of students' country of origin in 2002 in columns (1) – (3). Standard errors are clustered at country level. *, **, *** denote statistical significance at the 10%, 5% and 1% level.

TABLE A-5: The number of overseas students in China before and after the opening of a CI abroad in countries with CIs and when controlling for pre-treatment effects two years prior to an opening

	Dropping countries with no CIs			Controlling for $Open_{t+2}$		
	Total (1)	Long-term (2)	Short-term (3)	Total (4)	Long-term (5)	Short-term (6)
$Open_{t+2}$				−10.023	−6.754	−3.269
				(7.490)	(5.588)	(2.282)
$Open_{t+1}$	−0.273	−3.235	2.962	0.605	−2.293	2.898
	(10.940)	(8.297)	(3.132)	(10.877)	(8.250)	(3.125)
$Open_t$	10.022	1.852	8.170	9.867	1.857	8.010
	(15.205)	(10.571)	(5.029)	(14.992)	(10.355)	(5.022)
$Open_{t-1}$	17.353	9.662	7.691*	17.682	9.856	7.826*
	(18.440)	(14.387)	(4.394)	(17.186)	(13.335)	(4.238)
$Open_{t-2}$	4.338	0.777	3.561	3.782	0.154	3.629
	(15.506)	(12.584)	(4.172)	(15.430)	(12.483)	(4.062)
$Open_{t-3}$	23.580	14.799	8.781*	23.618	14.506	9.112*
	(16.146)	(12.468)	(4.862)	(15.752)	(12.102)	(4.773)
$Open_{t-4}$	23.937	13.930	10.007**	23.327	12.812	10.515**
	(16.366)	(12.769)	(4.641)	(14.958)	(11.661)	(4.205)
$Open_{t-5plus}$	79.896**	57.565**	22.331***	74.305**	51.790**	22.515***
	(33.306)	(28.089)	(6.734)	(30.584)	(25.463)	(6.410)
Other controls	Yes	Yes	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,489	1,489	1,489	2,250	2,250	2,250
# of countries	116	116	116	182	182	182

Notes: In columns (1) – (3), countries that have no institute during the observation period 2002 – 2014 are excluded from the estimation sample. Columns (4) – (6) show results when controlling for pre-treatment effects two years prior to an opening in the full sample. The dependent variables are the number of overseas students in total (columns (1) and (4)), for a long-term study stay (columns (2) and (5)), and for a short-term study stay (columns (3) and (6)) in China per 1,000,000 population aged 15 – 34 of the country of origin. The main independent variables are a set of dummies, $Open_{t-j}$. $Open_t$ takes value 1 if a new CI is established in the current year t. The dummy variables $Open_{t+1}$ and $Open_{t+2}$ bring forward an opening of a CI by one and two years respectively, and dummy variables $Open_{t-1}$ to $Open_{t-5plus}$ capture the response over time of overseas students to the opening of a new CI. Control variables considered in all regressions include country and year fixed effects, the logarithm of GDP per capita of students' country of origin ($lgdp_{-1}$), the logarithm of the absolute value of exports from students' country of origin to China ($lexp_{-1}$), and the logarithm of the absolute value of imports from China to students' country of origin ($limp_{-1}$). All control variables are measured one year prior to the measurement of outcome variables. Regressions are weighted by the 2002 population of students' country of origin. Standard errors are clustered at country level. *, **, *** denote statistical significance at the 10%, 5% and 1% level.

Appendix B Data on Confucius Institutes

Data sources

We collected information on CIs from the official web page of Hanban at: http://english.hanban.org in May 2017. CIs are listed and ordered by continents and countries. For each listed CI, we obtained information on the host country and the start running date (opening date) of a CI. In rare cases, information on the opening date of CIs was missing. We then checked the web page of the respective CIs as well as news items on the opening to supplement the missing information. Using this information, we recorded for each foreign country the number of CIs it hosted on the 31st of December of a year.

The *Hanban Annual Report*, also known as *Confucius Institute Annual Development Report*, also provides some information on the number of CIs. However, the first report was published only in 2006, so detailed information of CIs that opened before 2006 is not available. The 2006 report provides a list named as *Agreements Signed between Hanban and Overseas Institutions on the Establishment of CIs and CIs Formally Launched in 2006*, recording the overseas institutions that signed a letter of intent or agreement on the establishment of a CI and the overseas institutions that launched a CI in that year. Another list in the report provides the names of all CIs in 2006. However, this list includes also CIs merely in planning. Since 2008, annual reports provide no list of institutes that have been newly launched in a year and also no list of planned institutes, for which a letter of intent or agreement was signed. Since 2010, the reports only provide information on the number of CIs a country currently has. However, CI figures reported for countries are likely to include also CIs merely in planning. If a CI was closed, it is impossible to tell if new institutes opened in the same year and country, especially in countries that have a large number of CIs and that could have both openings and closures within the same year.

The above-mentioned two sources of information on CIs have been used in the existing literature (Akhtaruzzaman et al., 2017; Ghosh et al., 2017; Lien et al., 2012, 2014, 2017, 2018, 2019; Lien and Co, 2013; Lien and Oh, 2014). For the following reasons, we used the first data source, the web page of Hanban, to collect data. First, only these data source provides information on the actual opening date of CIs, which we need for our analyses, and such dates may vary across countries, time, and individual CIs from the date that mere agreements were signed for individual CI establishment.[25] Second, with information on the

[25]For instance, according to the annual report of 2006, Woosong University of South Korea signed an agreement in November 2006, and it was counted in the list of CIs in the 2006 report. However, the web page of Hanban shows that Woosong University started to operate the institute only on 30th April 2007 (see: http://www.hanban.org/confuciousinstitutes/node_6847_5.htm). In our data set, 2007 was recorded as the opening year of the CI in Woosong University.

exact opening date of CIs, we are able to analyze the development of overseas students before and after the opening of a CI in a foreign country and thereby gauge the potential importance of anticipation effects and pre-treatment effects more generally that could invalidate a causal interpretation of our estimated effects of CI openings on post-opening overseas student numbers.

Closures of CIs

It is possible, as noted, that some CIs closed during the observation period of our analysis. If such institutes closed before we accessed the Hanban data, they are not recorded in our data set. Closures of institutes, however, occurred only very rarely in our observation period (2002 – 2014) and in but few countries, including the US, Germany, France, Sweden, and Canada.[26]

[26]Information about closures of CIs was retrieved from: https://www.bbc.com/zhongwen/simp/world-45237598, https://www.bbc.com/zhongwen/simp/world/2013/04/130404_mcmaster_confucius_institute.shtml.

Appendix C Replicating Lin et al. (2016)

Lin et al. (2016) restrict the estimation sample to 40 countries that have the most overseas students in China. These countries include South Korea, Japan, Singapore, Ireland, Austria, Belgium, Denmark, Germany, France, the Netherlands, Portugal, Sweden, Spain, Italy, the UK, Switzerland, Canada, the US, Australia, New Zealand, Pakistan, Philippines, Malaysia, Thailand, Iran, India, Indonesia, Saudi Arabia, Vietnam, Russia, South Africa, Nigeria, Egypt, Mexico, Argentina, Brazil, Kazakhstan, Bangladesh, Ethiopia and Colombia. Lin et al. (2016) find a negative effect of CIs on overseas student numbers.

We replicated their estimation to gauge the reason for their divergent findings. The estimation sample is restricted to the above-mentioned 40 countries and the observation period is from 2004 to 2014, the same as in Lin et al. (2016). We employ the logarithm of the total number of overseas students in China as the outcome measure and consider as the main independent variable the absolute number of CIs one year prior to the measurement of students. As control variables, we use the logarithm of population, trade (exports plus imports), and GDP per capita as well as its squared term.[27] We exclude the measure of cultural distance between China and the 40 countries that is used in Lin et al. (2016). This culture variable originates from Qi et al. (2012), and consists of a time-invariant cultural distance measure from Kogut and Singh (1988) and the duration that diplomatic relations existed between a given country and China. However, no evidence is provided in Lin et al. (2016) that cultural distance decreases with the number of years of diplomatic relations. Time-invariant cultural distance can be controlled for by country fixed effects. It is also unclear why these two factors are put additively together, which assumes that the effect of these two factors on the number of overseas students is the same. Furthermore, the correlation coefficient between the number of CIs and the culture variable is very low (0.07), which indicates that the coefficient on the number of CIs may not change much if the culture variable is dropped from the regression model. Lin et al. (2016) employ a fixed effects model, but do not cluster standard errors at the country level.

Column (1) of Panel (A) in Table C-1 shows the result if we estimate the model without clustering, and we also find a significant negative effect of the number of CIs on the growth rate of overseas students, similar to the findings in Lin et al. (2016).[28] However, it is necessary

[27] We omitted the foreign direct investment between China and the 40 countries, since Lin et al. (2016) argue that it has no significant effect on the number of overseas students, and drop this covariate from the baseline specification.

[28] The number of observations in Lin et al. (2016) is 374. It is unclear why the sample is not balanced and why some observations are dropped. The size of the coefficient in Lin et al. (2016) is smaller than the one we have estimated. The different sample size may be a potential reason for the different coefficient in magnitude.

TABLE C-1: Results of replicating and comparing with Lin et al. (2016)

	Panel (A)			Panel (B)	
	log(students) (1)	log(students) (2)	abs. students (3)	std. students (4)	std. students (5)
CI^a_{-1}	−0.006**	−0.006*	163.507***		
	(0.003)	(0.003)	(20.941)		
CI				159.618**	284.032***
				(73.704)	(102.776)
$lpop_{-1}$	4.032***	4.032**	−9,648.805		
	(0.646)	(1.818)	(9,002.972)		
$ltrade_{-1}$	0.483***	0.483*	−1,282.727		
	(0.085)	(0.253)	(860.779)		
$lgdp_{-1}$	0.840***	0.840	7,348.118*	409.261	146.472*
	(0.311)	(0.910)	(3,967.769)	(266.330)	(73.525)
$lgdp^2_{-1}$	−0.045**	−0.045	−186.383		
	(0.019)	(0.050)	(239.255)		
$lexp_{-1}$				−133.374	−16.155
				(82.858)	(17.971)
$limp_{-1}$				−244.872*	−56.372*
				(143.504)	(31.539)
Country FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
Clustered S.E.	No	Yes	Yes	Yes	Yes
Weighting	No	No	No	No	Yes
Observations	440	440	440	440	440
# of countries	40	40	40	40	40

Notes: This table shows the effect of CI numbers on the number of overseas students in total when we use the estimation sample and the observation period applied in Lin et al. (2016). In panel (A), the main independent variable, CI^a_{-1}, is the absolute number of CIs of a CI-host country measured one year prior to the measurement of overseas students. The dependent variable in columns (1) and (2) is the logarithm of the total number of overseas students a country has in China and in column (3) the absolute number of overseas students in total. Control variables considered in this panel include country and year fixed effects, the logarithm of population of students' country of origin ($lpop_{-1}$), the logarithm of the absolute value of trade in students' country of origin ($ltrade_{-1}$), and the logarithm of GDP per capita of students' country of origin ($lgdp_{-1}$) as well as its squared term. All control variables are measured one year prior to the measurement of outcome variables. In panel (B), the main independent variable is the number of CIs per 1,000,000 population aged 15 – 34 of a CI-host country (CI). The dependent variable is the number of overseas students in total in China per 1,000,000 population aged 15 – 34 of the country of origin (columns (4) and (5)). Control variables considered in this panel include country and year fixed effects, the logarithm of GDP per capita of students' country of origin ($lgdp_{-1}$), the logarithm of the absolute value of exports from students' country of origin to China ($lexp_{-1}$), and the logarithm of the absolute value of imports from China to students' country of origin ($limp_{-1}$). All control variables are measured one year prior to the measurement of outcome variables. The regression in column (5) is weighted by the 2002 population of students' country of origin. Standard errors in columns (2) – (5) are clustered at country level. *, **, *** denote statistical significance at the 10%, 5% and 1% level.

to adjust standard errors, because observations across time in a particular country are likely to be correlated. When we cluster standard errors at the country level (column (2) of Table C-1), the estimated coefficient on CI^a_{-1} is only significant at the 10% significance level.[29] Moreover, changing the outcome measure to the absolute number of students and clustering standard errors at the country level, we find a significant positive effect of the number of CIs on the number of students (column (3) of Table C-1), which suggests that the functional form employed by Lin et al. (2016) may be wrong. The absolute number of students is a

[29]If we enlarge the observation period in this analysis to 2002 – 2014, the same as our baseline model, the significance of the estimated coefficient disappears.

linear function of the number of CIs, when other factors are held constant.[30]

In the end, we employ our estimation model in this study for the selected 40 countries, i.e. regressing the population share of overseas students on the population share of CIs and controlling for the logarithm of GDP per capita, exports to China and imports from China. We show the estimation results in Panel (B) of Table C-1. The regression in column (4) is unweighted and the regression in column (5) weighted by the 2002 population of students' country of origin. We still find a significant positive effect of the number of CIs on the number of overseas students in China. The size of the coefficient in column (5) is also close to the one in our baseline results (column (1) of Table 2).

[30]The negative effect of CI^{a}_{-1} on overseas student numbers in column (2) seems to indicate a concave effect. However, it might be driven by the US, a country with a very large number of CIs. If we drop the US from the estimation sample and use the specification in column (2), the coefficient is still negative, but much smaller and insignificant. Using the specification in column (3) and dropping the US from the sample, we still find a significant and positive effect of CI numbers on the number of overseas students in China.

Otto von Guericke University Magdeburg
Faculty of Economics and Management
P.O. Box 4120 | 39016 Magdeburg | Germany

Tel.: +49 (0) 3 91 / 67-1 85 84
Fax: +49 (0) 3 91 / 67-1 21 20

www.fww.ovgu.de/femm

ISSN 1615-4274