

RIETI Discussion Paper Series 22-E-085

Impact of Stay Abroad on Language Skill Development: Regression discontinuity evidence from Japanese university students

HIGUCHI, Yuki Sophia University

NAKAMURO, Makiko RIETI

ROEVER, Carsten University of Melbourne

SASAKI, Miyuki Waseda University

YASHIMA, Tomoko Kansai University



The Research Institute of Economy, Trade and Industry https://www.rieti.go.jp/en/

RIETI Discussion Paper Series 22-E-085 September 2022

Impact of stay abroad on language skill development: Regression discontinuity evidence from Japanese university students^{*}

> Yuki HIGUCHI (Sophia University) Makiko NAKAMURO (Keio University/RIETI) Carsten ROEVER (University of Melbourne) Miyuki SASAKI (Waseda University) Tomoko YASHIMA (Kansai University)

Abstract

The importance of English communication skills has been increasing with globalization, and the governments various countries have encouraged students to go abroad. However, the causal impact of staying abroad has been little investigated, particularly in non-European countries. This study adopts a regression discontinuity design to the Japanese government's flagship scholarship program for youths' studying abroad. We found that the scholarships significantly increased the probability of studying abroad by 40 points. By comparing the students placed close to the cutoff, we found that the scholarship increased English proficiency by 12% (or 0.42 standard deviation), measured by a multiple-choice test we originally developed. We also found that the scholarship significantly improved their international posture scores and the perceived communication competence in a foreign language, which are the two traits found as important determinants of future development in language ability in applied linguistics literature.

Keywords: study abroad, stay abroad, applied linguistics, RDD, impact evaluation, Japan JEL classification: F2, I2, J2

The RIETI Discussion Paper Series aims at widely disseminating research results in the form of professional papers, with the goal of stimulating lively discussion. The views expressed in the papers are solely those of the author(s), and neither represent those of the organization(s) to which the author(s) belong(s) nor the Research Institute of Economy, Trade and Industry.

^{*} Higuchi (corresponding author): Faculty of Economics, Sophia University (email: higuchi@sophia.ac.jp); Nakamuro: Faculty of Policy Management, Keio University (email: makikon@sfc.keio.ac.jp); Roever: School of Languages and Linguistics, University of Melbourne (email: carsten@unimelb.edu.au); Sasaki: Faculty of Education and Integrated Arts and Sciences, Waseda University (email: miyuki.sasaki@waseda.jp); Yashima: Faculty of Foreign Language Studies, Kansai University (email: yashima@kansai-u.ac.jp). This study is conducted as a part of the project entitled "Establishing Evidence-Based Policy Making in Japan," undertaken at the Research Institute of Economy, Trade and Industry (RIETI). We would like to thank Jun Goto, Sun Youn Lee, Hiroko Okudaira, Shinpei Sano, and Hiroshi Yoshida as well as seminar and conference participants at Tohoku University, Kansai Labor, 2021 Japanese Economic Association Spring Meeting, and the Discussion Paper seminar at RIETI for their helpful comments and suggestions. We also acknowledge the members of our target program for their active research collaboration, and we appreciate the survey respondents for providing their valuable information. This research was financially supported by the Tokyo Foundation for Policy Research, MEXT/JSPS KAKENHI Grant Number: 18H05314 and Grant-in-Aid for Research at Nagoya City University, where the first and fourth authors were affiliated with until March 2020. All errors are our own.

I. Introduction

With the progress of globalization, the number of students studying abroad dramatically increased in the past few decades. According to the United Nations Educational, Scientific and Cultural Organization (UNESCO, 2019), the number tripled between 1999 and 2018. The government in various countries encourages students to go abroad by disseminating information and providing financial support. The causal effect of studying abroad should be well understood to justify governmental support. Still, it has little empirically been investigated because of the self-selection problem and the consequent difficulties in the empirical analysis.

While Waibel et al. (2017) conducted a systematic review of 65 studies analyzing the "impacts" of studying abroad, only six adopted causal inference methods. Other studies mostly relied on beforeafter comparison and/or with-without (i.e., study abroad v.s. stay home) comparison, both of which are subject to problems for causal identification. The six studies are all conducted in European countries. While there are a few more recent studies conducted outside Europe, the scarcity of the existing literature warrants more investigations in non-European countries, where the language distance from English is long, and the expected impacts of studying abroad are likely different.

To fill this gap, we adopted the regression discontinuity design (RDD) to evaluate the impacts of the Japanese government's flagship scholarship for university students, including graduate students, to stay abroad¹. We found that the scholarship increased the probability of staying abroad by 40 percentage points. This means that almost all the scholarship winners went abroad, and 60% of the losers also went abroad, using their own funding or another scholarship. By comparing the students near the cutoff, we found that the receipt of the scholarship increased English proficiency by 12% (or 0.46 standard deviation). To measure English proficiency, we have developed a reasonably valid

¹ The term "study abroad" mostly describes learners going to the target language environment specifically to study the target language. The broader concept of "stay abroad" includes learners studying subject content in the target language environment through the target language or a lingua franca or learners working temporarily in another country and communicating in the local language or a lingua franca. We adopt the term "stay abroad" since our target program includes a wide range of international experiences.

multiple-choice test consisting of 12 questions that can be completed in 12 minutes for the targeted population (Roever et al., 2022).

We also conducted our original survey to measure the impacts on their international posture and perceived competence in a foreign language. These two traits are found as important determinants of future language development in applied linguistics literature (e.g., MacIntyre and Charos, 1996; Yashima, 2002; Yashima and Zenuk-Nishide, 2008). Our data shows that the scholarship improved the international posture score by 7% and the perceived communication competence score by 25%.

Although we found significant impacts on English proficiency as well as on the important traits associated with language development, the impacts are modest compared to the existing studies that did not adopt causal inference methods. In particular, <u>Yokota et al. (2018)</u> conducted a large-scale survey of more than 6,000 respondents. They found that, compared to those who have never studied abroad, those who studied abroad for at least three months have greater skills in all the 18 dimensions they measured. Our finding suggests that such a simple with-without comparison is most likely to overstate the impact of stays abroad because of the issue of self-selection.

The remainder of this paper is organized as follows. Section II reviews the existing literature on the impacts of study/stay abroad. Section III describes the institutional setting, and Section IV explains our data. Section V presents the empirical specification and results, and Section VI concludes with a few remarks.

II. Literature

There has been accumulated research on the effect of studying or staying abroad on language proficiency in the field of applied linguistics. <u>Freed (1995)</u> is one of the early edited books on this topic, which compiles several case studies from different countries. According to a recent meta-analysis conducted by <u>Tullock and Ortega (2017) and Yang (2016)</u>, a positive association is generally observed between study abroad experience and language proficiency gain.

In the economics literature, while some studies found that the acquisition of language proficiency has a high return in the labor market (e.g., <u>Bleakley and Chin, 2004; Sorrenti, 2017</u>), the impact of studying or staying abroad is little investigated. Only some recent papers try to address the problem by using causal inference methods to examine the impacts on language development as well as labor market outcomes. <u>Parey and Waldinger (2011)</u> is a pioneering work in economics. They used department- and year-level variations in the enrollment to the European community action scheme for the mobility of university students (the ERASMUS program) as an instrument for studying abroad for German university students. They found that the experience of studying abroad made them interested in a foreign culture and encouraged them to work overseas. <u>Di Pietro (2012, 2015)</u> used the same instrument variable approach for Italian university students and also found that the students studying abroad experience were significantly more likely to work abroad.

<u>Oosterbeek and Webbink (2011)</u> adopted a similar identification strategy as ours. They used a regression discontinuity design for a government-funded scholarship for Dutch university students. They found that the students with study abroad experience were more likely to live abroad at the time of the survey. <u>Waibel et al. (2018) and Liwiński (2019)</u> used the propensity score matching method for German and Polish university students, respectively, and found that studying abroad increased the employment probability. Since labor mobility is high among European countries, these studies mostly focused on labor market outcomes, particularly international migration.

<u>De Poli et al. (2018)</u> investigated non-labor market outcomes. They conducted a randomized controlled trial for a one-month short study abroad program in an Italian university, finding that the treated students scored significantly higher in an English proficiency test. They also found that the treated students had stronger personal traits, including willingness to communicate, social orientation, adaptability, and openness. While the studies mentioned above are all conducted in European countries, the impact of studying abroad is likely to be different in non-European countries.

Two sets of notable studies conducted outside Europe are Kawata and Nishitani (2017), and

<u>Kashima and Kato (2021) and Kato and Suzuki (2019)</u>. <u>Kawata and Nishitani (2017)</u> used differencesin-differences estimation for a short study-abroad program for university students and found that the program improved the TOEIC score by 7%. <u>Kato and Suzuki (2019)</u> conducted a randomized controlled trial in a similar short study-abroad program for university students. They found that the program significantly increased the probability that the same student would go for a longer studyabroad program. <u>Kashima and Kato (2021)</u> used the same data and found that the treated students were more likely to be employed by companies listed on the first section of the Tokyo Stock Exchange, i.e., well-known large companies. While these studies analyzed the impact of short-term study abroad programs among Japanese university students, we estimate the impact of longer-term study abroad on language skill development.

III. Setting

As a part of the "Action Plan of the Growth Strategy" initiated by the Prime Minister of Japan, the Ministry of Education, Culture, Sports, Science, and Technology (MEXT) launched its flagship program to provide scholarships for Japanese youths to stay abroad in 2014. It is a seven-year program aiming to financially support a total of 10,000 students in high schools, universities, and graduate schools². Since the undergraduate and graduate students were recruited and treated equally in the selection process, we combined the undergraduate and graduate students and called them "university" students for readability. We target university students in our analysis because the typical study duration is only a month for high school students, while it is nine months for university students.

The unique feature of this scholarship program is that private companies and individuals fully sponsor it. The 240 companies, mostly leading companies in Japan, donated 200 million USD to budget the seven-year program. Partly because of this, the program provides flexible scholarships, and the

 $^{^2}$ The program was planned to be concluded in 2020, but it has been extended to 2022 due to COVID-19. In August 2022, the further program extension to 2027 was officially announced because of its established reputation.

students choose destination countries and durations of stay (four weeks to two years for university students and two weeks to one year for high school students). While the majority of students stayed in English-speaking countries, approximately about a third chose non-English-speaking countries. Furthermore, they can design their own study-abroad plans, including exchange programs, study in schools, internships, and volunteer work. The selection criteria do not depend on the English (or other languages) proficiency at the time of application, resulting in the English skills before going abroad varied considerably among our sample. Rather, the main criteria for selection are more on students' motivation, such as curiosity and passion.

The selection was conducted twice a year for the university students, and the selection process consisted of two stages. The first stage narrowed down applicants based on written applications, including resumes and study-abroad plans. The application documents were reviewed by selection committees consisting of two university faculty members and one person from the sponsoring companies. If an applicant passed the document screening, s/he proceeded to the interview. The individual and group interviews were conducted, and each was independently judged by a person from a sponsoring company. The final decision was made based on the weighted average of the evaluation scores for the document screening and the interviews.

We have worked with MEXT and collected data from our target batch. In this cohort, 1,496 students submitted the application documents, and 709 passed the document screening and took the interview; these 709 students consist of our sample students. The numbers are reported in Table 1. Some students who passed the document screening but declined to take the interview (and thus, the scholarship) are out of the scope of our study. The 709 students included the 451 students who eventually passed the selection and won the scholarship (we called them the "treated" group) and the 258 students who did not pass the selection (we called them the "control" group). Once students are selected for the program, they receive the tuition fee, living expenses (1,150 to 1,540 USD per month), and transportation fees (1,440 to 2,410 USD). This selection process creates the cutoff of applicants to

receive this financial support, allowing us to exploit the discontinuities.

There are four programs in the target scholarship program: a top-level university course, natural science course, an emerging country course, and a variety of talent course. The first mainly targets students who want to study at a top-level university, and the second mainly targets students majoring in natural science. The third mainly targets students who want to study the local language or other subjects, do volunteer activities, or do internships in emerging countries. The last targets students who want to join various art and sports activities. The selection process was the same for all the courses, and students were evaluated together. However, the minimum score for winning the scholarship differs depending on the allocated seats for each course and the number of applicants. The selection is basically based on the weighted average of document screening and two interviews (individual and group). However, some adjustments were made based on the extra remarks from the interviewers as well as the household income. For example, some students whose score is slightly below the cutoff could win the scholarship if they received bonus points from any interviewer or if their household income is below a certain threshold. Further, there is a university-level adjustment to rescue applicants from a university with a certain number of applicants but without winners based on the selection score. With such adjustments, we adopt a fuzzy RDD design, using the original weighted average score as a running variable.

IV. Data

Data collection

We targeted the 709 students who passed the document screening and conducted a baseline survey at the time of the second stage interview. The survey was conducted in January 2018. We explicitly explained to the sample students that the study was not a part of the selection process and that their answers (including refusal) would never influence the selection results. Probably because we conducted the baseline survey onsite, the response rate was as high as 94%, although we did not

provide any incentive to complete the survey. The timeline of the survey is presented in Table 1. We collected the baseline data from all the students who took the interview. Our identification strategy is to compare those who are above the threshold to receive the scholarship (the treated group) and those who are below it and fail to receive the scholarship (the control group).

After the baseline survey, we conducted two rounds of follow-up surveys. The first round was conducted in October 2019. We sent an email invitation for the online survey to the treated and the control students with the reward of 9 USD worth of an Amazon gift card upon completing the survey. The response rate was 52% (we will discuss the issue of attrition more in the following subsection). The second follow-up survey was conducted in June 2020. To increase the response rate, we increased the incentive amount to an 18 USD worth of Amazon gift card. The response rate was 44%.

Basically the same questionnaires were used in the baseline and follow-up surveys. There are two notable differences. First, we collected the parental socio-economic status only in the baseline survey because it is essentially time-invariant. Second, the English test was conducted only in the follow-up surveys. In our baseline survey, we administered a test called Versant to measure English reading, listening, and speaking skills, which is based on a phone call and can be completed in 20 minutes. Yet, only half of the survey respondents took the test, presumably because they hesitated to download the app to take the test by making a phone call, which was new to most of the sample students. Given the low baseline response rate, we decided not to use the Versant scores.

Instead, we developed a test consisting of 12 questions that can be completed in 12 mins, all of which are multiple-choice questions with four answering options, to measure English proficiency. We (Roever et al., 2022) conducted a study to ascertain the validity of the test targeting a sample population similar to those targeted in this study. The results of the study were satisfactory (e.g., the correlation between the text and TOEFL iTP was .59), and we renamed the test as "Pragmatic test (PT)" and used the test for the subsequent surveys. The test was conducted at the end of the two rounds of follow-up surveys but not at the baseline survey. The difficulty of the test in each round was adjusted to be

comparable, with some questions being the same to check the consistency. Since we did not give respondents feedback with correct answers, it is unlikely that the students remembered the question in the second follow-up survey. Two examples are below (the underlined option is the correct answer).

Carrie has done some shopping at a grocery store. The man at the cash register has just finished packing her groceries and gives her the bags. What would the man probably say?

- 1. "There they are."
- 2. "I kept you waiting."
- 3. "Here you go."
- 4. "Please."

Tom ordered a meal in a restaurant and the waitress just brought it. She asks him if he wants to order additional items. What would the waitress probably say?

- 1. "Would you like anything extra?"
- 2. "Did you get everything ready?"
- 3. "What can I bring you?"
- 4. "Can I get you anything else?"

We counted the number of correct answers and used the score as our outcome measure for English proficiency.

To measure the impacts of staying abroad on attitude, we collected data on international posture and perceived competence in a second language. These are two of the few quantifiable constructs previously measured and used in a valid and reliable manner in the applied linguistics studies on studying abroad (for the international posture, see Botes et al., 2020; Lee, 2018; Yashima et al., 2004 and for the perceived competence, see MacIntyre and Charos, 1996; Yashima and Zenuk-Nishide, <u>2008</u>) These two constructs are related to the aim of the targeted scholarship programs to cultivating global workforce and can be measured by questionnaire items only. In addition, they are measured in a target-language-neutral manner because scholarship recipients were expected to travel to a country of their choice, including non-English-speaking countries.

To measure the international posture, we originally selected the 20 items used in <u>Yashima (2009)</u> and pretested them in our pilot survey targeting a previous batch of the same scholarship program. We found that many items had ceiling effects and decided to use the following four items to save survey time: (i) I am interested in international issues, (ii) I want to participate in a volunteer activity to help foreigners living in my surrounding community in Japan, (iii) I would like to talk to international students if there are any in my school in Japan, and (iv) I often read and watch news about foreign countries. Respondents were asked to choose their position for each item from 1 (not at all) to 6 (very much) on a Likert scale. We use the average responses to these four items as our measure of international posture.

To measure perceived competence in a second language, we asked how well respondents felt they could communicate from 0% (not at all) to 100% (very well) in the following four situations: (i) Speak in a group of friends, (ii) Speak in a meeting you attend for the first time, (iii) Speak to a person you meet for the first time, and (iv) Make a presentation in front of a large group of strangers. The competence was asked for the language they would use in a country they would visit (and thus, it is not necessarily English). The original study of McIntyre and Charos (1996) used 12 items, but we chose these four to save survey time, following Yashima (2002; 2009). We use the average responses to these four items as our measure of perceived competence in a second language.

Attrition

Since the response rate for the follow-up surveys is not high, we estimated the attrition regressions and report the results in Table 2. The outcome variable is a dummy variable taking the

value of one if a sample student responded to each survey. The gender dummy, the grade at the time of application, the dummy variable indicating whether the household income is above the threshold³, and the English proficiency presented in the Common European Framework of Reference for Languages (CEFR) are included as explanatory variables. These data were collected in the application documents and available for all the sample students except for the English proficiency. The submission of the test of English proficiency was optional because the scholarship program explicitly announces that English proficiency is not one of the judging criteria. We also control the treatment dummy, which takes the value of one if a student scored higher than the threshold as well as the standardized score for the selection. Since the threshold is different for each course category, we standardize the score by deducting the mean and dividing by the standard deviation (SD). In other words, our running variable is presented in a standardized z-score, and 0 means the minimum passing score. In column 1, the results show that only the dummy variable for not reporting English score is marginally significant. This indicates that most of these characteristics had limited explanatory power on the baseline attrition, and we assume that the attrition is not a serious problem for the baseline survey.

In columns 2 to 5, the treatment variable is statistically significant. This means that the students who won the scholarship are more likely to have responded to our follow-up surveys. In columns 2 and 4, we controlled for the same variables as in column 1, that is, the information provided in the application form, and thus, is available to all the sample students. In columns 3 and 5, we added various variables collected in the baseline survey. Hence, the second analysis is performed only among the sample students who responded to the baseline survey. The results show that male and older students were less likely to respond to the follow-up surveys, but no other coefficient is statistically different from zero. To the extent that the explanatory power of a rich set of observable characteristics is limited, we assume that the degree to which the unobservable factors influence attrition is limited.

³ The threshold depends on the number of siblings, whether a student is enrolled in a national/public or private university (the tuition fee is higher in a private university), and whether s/he is living with the parent. For a household with one child who goes to a national/public university and lives together, the threshold is about 90,000 USD.

Furthermore, in the second follow-up survey, we recorded the email addresses to which the invitation for the survey was sent but returned to the program secretariat. 10% of the emails were returned, and only 90% of the sample students received the invitation for the second follow-up survey. Columns 6 and 7 show the attrition regression results only among those who receive the email. Importantly, the coefficient of the treatment dummy became smaller and statistically insignificant, indicating that the treated students were no more likely to respond to the survey. Hence, we assume that the attrition pattern is not systematically different between the two groups if we condition on having received the email.

Descriptive analyses

To explain the collected data, Table 3 presents the summary statistics of our main outcome variables. The difference seems smaller, but for all the three indicators we have, the score is higher for the treated students, even at the baseline. Although more than half of the applicants were screened out in the document selection and our sample students are homogeneous to some extent, there may be notable differences between those who eventually passed the interview and those who failed. We adopt the local approach to compare the students closer to the cutoff to estimate the causal impacts of the scholarship.

Before presenting the local estimation, Figure 1 illustrates the first-stage impact of the scholarship among our sample students. It plots the standardized score (i.e., the running variable) on the horizontal axis and the scholarship acquisition (in Panel A), the study abroad probability for four weeks or longer (in Panel B), and the study abroad duration (in Panel C) in the vertical axis. Panel A shows that there is a jump at the cutoff score, but many students below the cutoff also received the scholarship because of the adjustment. Therefore, we adopt fuzzy RDD in the estimation. Panel B shows that the students above the cutoff are about 40 points more likely to study abroad. This suggests that the scholarship indeed promoted university students to go abroad, while 60% of the students went

abroad, using other scholarships or self-financing. Panel C shows a jump in the study abroad months at the cutoff. Much of this jump is explained by the study abroad probability (i.e., the extensive margin). If we focus only on the student who stayed abroad, its length has only a small jump at the cutoff.

V. Results

Empirical strategy

By taking advantage of the selection process, we adopt RDD. Based on the methods developed by <u>Imbens and Kalyanaraman (2012)</u>, the optimal bandwidth is computed as 0.903. If we use an alternative method developed by <u>Calonico et al. (2014</u>), the optimal bandwidth is 0.747, which is not largely different from 0.903. For readability, we use the 0.903 bandwidth in our main analysis, and we report the result using the 0.747 as well as other bandwidths as sensitivity checks. The 0.903 means that we limited our attention to the students whose score was in the plus and minus 0.9 standard deviation range from the cutoff.

We estimated both OLS and two-stage instrumental variable (IV) models. In the OLS model, we used the dummy variable taking the value of one if the score is above the threshold. The IV model is used to adjust for the fuzzy nature of our setting. We used the same dummy variable as an instrument for another dummy variable, which takes the value of one if a student won the scholarship. The estimated coefficient in the IV model indicates the local average treatment effect (LATE) of receiving the scholarship for a student with a score equal to the cutoff. Following the lead of <u>Gelman and Imbens</u> (2019) and Imbens and Lemieux (2008), we used the local liner regression and controlled the running variable on either side of the cutoff (and not their polynomials).

We have three main outcomes; the pragmatic test, the international posture, and the perceived competence. The pragmatic test was administered only in the follow-up surveys, but the latter two variables were collected in the baseline survey too. To improve the precision of the estimation, we follow <u>McKenzie (2012)</u> and control the baseline levels on the right-hand side for these two variables.

Validity

We report the results of two checks on the internal validity of our empirical strategy. First, we conducted the <u>McCrary (2008)</u> test to check whether the manipulation occurs at the threshold. Figure 2 shows the smooth distribution without any bunching at the cutoff. The *p*-value for the existence of bunching is insignificant. Second, based on the same right-hand-side specification as explained above, we checked whether the baseline characteristics, including gender, grade, the household income dummy, baseline English proficiency, and parental socio-economic status, are balanced within the bandwidth. Table 4 shows that none of the treatment coefficients is significant, illustrating that the treated and the control students are comparable in these observable characteristics. All of these variables were controlled both in our OLS and IV estimation to increase its precision.

Results

We already show the relationship between the running variable and the stay-abroad outcomes in Figure 1, but Table 5 shows the estimated relationship among the students whose score is within our optimal bandwidth. Columns 1 and 3 report the OLS results, where the key variable of interest is whether the selection score of the students is above the threshold. Columns 2 and 4 report the IV results, whether the key variable is whether the student won the scholarship. OLS results show that the students above the cutoff are 40 points more likely to have stayed abroad and stayed abroad for 5.2 months longer than those below the cutoff. The coefficients almost doubled in the IV estimation. These results indicate that the presence of the scholarship indeed encouraged university students to stay abroad.

Table 6 shows the main estimation results. We report the OLS results in the odd number columns and the IV results in the even number columns. We conducted two rounds of follow-up surveys and report the pooled results in columns 1 and 2. When a student responded in both rounds of survey, we treat each response as an independent observation to increase the statistical power. In columns 3 and 4, we report the results using only the first follow-up survey. In columns 5 and 6, we report the results using only the second follow-up survey. As discussed in section 3, the attrition in the second follow-up survey is less likely to influence the results because the attrition is almost at random, conditional on receiving the email invitation to participate in the survey. While we need to consider such potentially different attrition patterns, the difference between columns 3 and 4 and columns 5 and 6 suggests the trajectory of lasting impacts.

The pooled data for the pragmatic test shows a significant coefficient of 0.83 for the OLS estimation and 1.66 for the IV estimation, respectively. The control mean is 7.00 and the standard deviation is 1.98, so the ITT estimate means a 12% increase or a 0.42 standard deviation increase in the PT test. The LATE estimate is even greater and indicates a 24% (or a 0.84 standard deviation) increase. Hence, the scholarship program indeed helped the students develop their English skills. Panels B and C show that the scholarship significantly improved the international posture and the perceived competence in a second language. Compared to the control means and standard deviations, the coefficient sizes are non-negligible.

If we look at the breakup of the impacts in the two follow-up surveys, the coefficient of the pragmatic test is only significant in columns 5 and 6 (although it is still positive in columns 3 and 4). In contrast, the coefficient of the international posture is only significant in columns 3 and 4. A possible interpretation is that it takes time and requires the students' effort to improve language skills to the extent that a significant difference is detected in statistical analysis. And the improved attitude encouraged the sample students to continue improving or investing in their language skills, and a significant improvement was observed later.

As a robustness check, Figure 3 presents the sensitivity analysis using different bandwidths. This analysis is based on the pooled data and used OLS estimation; thus, it is comparable to column 1 in Table 6. We have used the Imbens-Kalyanaraman bandwidth of 0.903 throughout this paper, and Figure 3 covers the Calonico et al. bandwidth of 0.747. The estimated coefficient is unstable with the smaller

bandwidths because the effective sample size gets smaller. In addition, the fact that several students who were just below the cutoff won the scholarship due to the adjustment particularly influences the estimation with narrow bandwidths. Yet, a general pattern is that the 95% confidence interval does not cross zero in Panel A for the pragmatic test and Panel C for the perceived competence when the bandwidth is greater than the optimal ones. The confidence intervals cross zero in Panel B, but the coefficient is stable and consistently above zero with the bandwidth above the optimal ones. Hence, we conclude that our results are not too sensitive for the bandwidth selection.

Remarks on non-cognitive skills

Lee (2019) found that university students with study abroad experience had greater noncognitive skills, such as aggressiveness, flexibility, and teamwork attitude. We hypothesized that the experience abroad might improve non-cognitive skills and collected information on Big 5 traits, GRIT, and self-efficacy measures. We estimate the impact of the scholarship on these non-cognitive measures using the same specifications described above, but almost all the coefficients are statistically insignificant. The insignificance may suggest that these non-cognitive skills are difficult to develop and are not easily developed by staying abroad experience of several months. Alternatively, Lee (2019)'s analysis is based on the comparison of students with and without studying abroad experience, and the selection effect largely explains the difference. Another possible explanation is that the students' reference points had been changed. Since the students or anyone who chose to stay abroad are likely to be different from their peers in the domestic networks, the way how they responded to the selfreporting questions on non-cognitive measures might have been altered.

VI. Conclusion

In this paper, we adopted RDD for the Japanese governmental flagship program for supporting youth staying abroad to evaluate its causal effect. We find that the scholarship significantly improved

their language skill development. This finding supports the success of the program but with a few remarks. First, to discuss the cost-effectiveness of governmental support of this type, we need to evaluate how the improved language skills will translate into increased income or productivity. Yet, the computation of the return on language skill is not easy and requires many assumptions. Second, there could be an impact of the program on aspects we cannot measure. In particular, the staying abroad experience may expand the students' network and exposure to more diverse people, providing them with a life-long impact. Hence, we cannot fully answer whether the scholarship to stay abroad is justifiable for the taxpayers.

Still, we provide causal evidence that the scholarship has significantly improved English proficiency as well as the attitudinal aspects of the sample students. Given that the international posture and perceived competence were found as important determinants of future development in language ability, a lasting and larger impact could be observed, possibly in labor or marriage markets, in the longer run.

Reference

- Bleakley, H. and Chin, A. 2004. Language Skills and Earnings: Evidence from Childhood Immigrants, *The Review* of *Economics and Statistics*, vol. 86, no. 2, 481–96
- Botes, E., Gottschling, J., Stadler, M., and Greiff, S. 2020. A systematic narrative review of International Posture: What is known and what still needs to be uncovered, *System*, vol. 90, 102232
- Calonico, S., Cattaneo, M. D., and Titiunik, R. 2014. Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs, *Econometrica*, vol. 82, no. 6, 2295–2326
- De Poli, S., Vergolini, L., and Zanini, N. 2018. The impact of a study abroad programme on learning abilities and personality traits: evidence from a randomization, *Applied Economics Letters*, vol. 25, no. 8, 562–66
- Di Pietro, G. 2012. Does studying abroad cause international labor mobility? Evidence from Italy, *Economics Letters*, vol. 117, no. 3, 632–35
- Di Pietro, G. 2015. Do Study Abroad Programs Enhance the Employability of Graduates?, *Education Finance and Policy*, vol. 10, no. 2, 223–43
- Freed, B. F. 1995. Second Language Acquisition in a Study Abroad Context, John Benjamins Publishing Company
- Gelman, A. and Imbens, G. 2019. Why High-Order Polynomials Should Not Be Used in Regression Discontinuity Designs, *Journal of Business & Economic Statistics*, vol. 37, no. 3, 447–56
- Imbens, G. and Kalyanaraman, K. 2012. Optimal Bandwidth Choice for the Regression Discontinuity Estimator, *The Review of Economic Studies*, vol. 79, no. 3, 933–59
- Imbens, G. W. and Lemieux, T. 2008. Regression discontinuity designs: A guide to practice, *Journal of Econometrics*, vol. 142, no. 2, 615–35
- Kashima, R. and Kato, M. 2021. Short-term Study Abroad Effect on Learning Outcomes and Initial Career: Evidence from Random Assignment Data from a Japanese University, *mimeo*
- Kato, M. and Suzuki, K. 2019. Effective or Self-Selective: Random Assignment Demonstrates Short-Term Study Abroad Effectively Encourages Further Study Abroad, *Journal of Studies in International Education*, vol. 23, no. 4, 411–28
- Kawata, K. and Nishitani, H. 2017. Tanki ryugaku puroguramu ga gogakuryoku toutatsudo ni ataeru eikyou ni tsuite: Hiroshima daigaku START puroguramu no jirei kara, *mimeo*
- Lee, J. H. 2018. The effects of short-term study abroad on L2 anxiety, international posture, and L2 willingness to

communicate, Journal of Multilingual and Multicultural Development, vol. 39, no. 8, 703-14

- Lee, S. Y. 2019. Kaigai ryugaku no kouka ni kansuru jissyou kenkyuu: Hininchi nouryoku to roudou shijyou no seika wo chushin ni, *Kokusaigaku kenkyu*, vol. 54, 1–28
- Liwiński, J. 2019. Does studying abroad enhance employability?, *Economics of Transition and Institutional Change*, vol. 27, no. 2, 409–23
- MacIntyre, P. D. and Charos, C. 1996. Personality, Attitudes, and Affect as Predictors of Second Language Communication, *Journal of Language and Social Psychology*, vol. 15, no. 1, 3–26
- McCrary, J. 2008. Manipulation of the running variable in the regression discontinuity design: A density test, *Journal* of *Econometrics*, vol. 142, no. 2, 698–714
- McKenzie, D. 2012. Beyond baseline and follow-up: The case for more T in experiments, *Journal of Development Economics*, vol. 99, no. 2, 210–21
- Oosterbeek, H. and Webbink, D. 2011. Does Studying Abroad Induce a Brain Drain?: Does studying abroad induce a brain drain?, *Economica*, vol. 78, no. 310, 347–66
- Roever, C., Higuchi, Y., Sasaki, M., Yashima, T., and Nakamuro, M. 2022. Validating a test of L2 routine formulae to detect pragmatics learning in stay abroad, *Applied Pragmatics*, Advance Access published 4 March 2022: doi:10.1075/ap.20015.roe
- Sorrenti, G. 2017. The Spanish or the German apartment? Study abroad and the acquisition of permanent skills, *Economics of Education Review*, vol. 60, 142–58
- Tullock, B. and Ortega, L. 2017. Fluency and multilingualism in study abroad: Lessons from a scoping review, *System*, vol. 71, 7–21
- Waibel, S., Petzold, K., and Rüger, H. 2018. Occupational status benefits of studying abroad and the role of occupational specificity - A propensity score matching approach, *Social Science Research*, vol. 74, 45–61
- Waibel, S., Rüger, H., Ette, A., and Sauer, L. 2017. Career consequences of transnational educational mobility: A systematic literature review, *Educational Research Review*, vol. 20, 81–98
- Yang, J.-S. 2016. The Effectiveness of Study-Abroad on Second Language Learning: A Meta-Analysis, Canadian Modern Language Review, vol. 72, no. 1, 66–94
- Yashima, T. 2002. Willingness to Communicate in a Second Language: The Japanese EFL Context, *The Modern Language Journal*, vol. 86, no. 1, 54–66

- Yashima, T. 2009. International posture and the ideal L2 self in the Japanese EFL context., pp. 144–63, in Dörnyei,Z. and Ushioda, E. (eds.), *Motivation, language identity and the L2 self*, Multilingual Matters
- Yashima, T. and Zenuk-Nishide, L. 2008. The impact of learning contexts on proficiency, attitudes, and L2 communication: Creating an imagined international community, *System*, vol. 36, no. 4, 566–85
- Yashima, T., Zenuk-Nishide, L., and Shimizu, K. 2004. The Influence of Attitudes and Affect on Willingness to Communicate and Second Language Communication: Language Learning, *Language Learning*, vol. 54, no. 1, 119–52
- Yokota, M., Ota, H., and Shimmi, Y. 2018. *Kaigai ryugaku ga kyaria to jinsei ni ataeru inpakuto: Daikibo chousa ni yoru ryugaku no kouka sokutei*, Gakubunsha

Table 1: Sample and timeline

Applicants	1496
Those who passed the document screening	709
(= Study sample)	(T=451, C=258)
Baseline survey	668
(January 2018)	(T=420, C=248)
1st follow-up survey	372
(October 2019)	(T=262, C=110)
2nd follow-up survey	312
(June 2020)	(T=213, C=99)

Note: T stands for the treated group, i.e., the students who won the scholarship, and C stands for the control group, i.e., the students who passed the document screening but not the interview, failing to receive the scholarship.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Deceline	1 st £_11		and fallow		2 nd follow	-up survey
	Baseline	15, 101	low-up	2 nd Iollow	-up survey	(received th	e email only)
= 1 if passed the selection	-0.0065	0.22***	0.23***	0.13**	0.12*	0.067	0.053
	(0.026)	(0.061)	(0.064)	(0.061)	(0.064)	(0.067)	(0.071)
Selection score	-0.013	-0.032	-0.027	-0.014	0.0020	-0.0050	0.014
	(0.013)	(0.030)	(0.032)	(0.030)	(0.031)	(0.032)	(0.035)
=1 if male	0.0096	-0.12***	-0.11***	-0.090**	-0.075*	-0.12***	-0.11**
[0-1]	(0.018)	(0.038)	(0.041)	(0.038)	(0.042)	(0.041)	(0.044)
Grade at baseline	-0.0095	-0.022*	-0.021*	-0.025**	-0.027**	-0.024**	-0.027**
[9–21]	(0.0059)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.013)
=1 if HH income high	0.018	-0.047	-0.043	-0.037	-0.0094	-0.043	-0.0031
[0-1]	(0.025)	(0.067)	(0.070)	(0.068)	(0.071)	(0.071)	(0.073)
Baseline English CEFR level	0.0030	0.027	0.013	0.010	0.0075	-0.0079	-0.0097
[1-6]	(0.012)	(0.025)	(0.028)	(0.025)	(0.028)	(0.027)	(0.030)
=1 if no baseline English level	-0.063*	0.00013	-0.0059	0.020	0.028	0.034	0.054
	(0.034)	(0.055)	(0.058)	(0.056)	(0.060)	(0.059)	(0.065)
International posture			-0.0056		0.0070		0.0064
[1-6]			(0.026)		(0.026)		(0.028)
Perceived competence in L2			0.00080		-0.00097		-0.0012
[0-100]			(0.00092)		(0.00093)		(0.00098)
Father educational attainment			-0.012		-0.019		-0.020
[1-6]			(0.016)		(0.016)		(0.017)
Mother educational attainment			0.0018		0.012		0.013
[1-6]			(0.018)		(0.018)		(0.019)
Number of books at home			0.013		0.013		0.0054
[1-6]			(0.014)		(0.014)		(0.015)
N	709	709	646	709	646	637	579
R squared	0.020	0.049	0.054	0.027	0.031	0.027	0.032

Table 2: Attrition regression (=1 if responded to the survey)

Note: Estimated coefficients are reported. ***, **, and * indicate the 1%, 5%, and 10% levels of statistical significance, respectively. Numbers in parentheses are t-statistics based on heteroscedasticity-robust standard errors.

	(1)	(2)	(3)	(4)
			Perceived	competence
Outcome	РТ	Intl posture	L2	Japanese
Baseline				
T(N = 420)	No data	4.65	56.3	85.2
1(N-420)	No data	(0.79)	(22.9)	(12.7)
C(N - 249)	N- 1-4-	4.54	51.4	84.7
C(N = 248)	No data	(0.78)	(21.7)	(13.1)
1 st follow-up				
T(N = 2(2))	7.37	4.82	61.8	85.1
1 (N = 202)	(1.81)	(0.76)	(18.4)	(13.7)
C(N - 110)	7.08	4.67	57.8	84.0
C(N = 110)	(1.80)	$\begin{array}{c c c c c c c c c c c c c c c c c c c $	(12.6)	
2 nd follow-up				
- T ()I ()I)	8.11	4.86	59.7	84.9
1 (N = 213)	(1) (2) $PT Intl posture$ No data (0.79) A.54 (0.79) A.54 (0.78) $7.37 4.82$ (1.81) (0.76) 7.08 A.67 (1.80) (0.89) $8.11 4.86$ (2.10) (0.73) 7.18 A.75 (2.38) (0.91)	(19.0)	(13.4)	
$C(\mathbf{M} = 00)$	7.18	4.75	54.0	84.2
C(N = 99)	(2.38)	(0.91)	(21.5)	(15.6)

Table 3: Global comparison

Note: Means are reported, and standard deviations are reported in parentheses.

Table 4: Balance check

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
			Application d	ata			Baseline survey	
	=1 if male	Grade at	=1 if HH	Baseline	=1 if no	Father	Mother	Number of
		baseline	income	English	baseline	educational	educational	hooks at home
			high	CEFR level	English level	attainment	attainment	books at nome
= 1 if score is above the cutoff	-0.17	0.20	-0.035	-0.030	-0.0017	0.32	0.42	-0.22
	(0.12)	(0.39)	(0.079)	(0.20)	(0.079)	(0.36)	(0.35)	(0.38)
Selection score	0.47**	0.45	0.23**	-0.0070	0.13	0.017	-0.43	0.040
	(0.22)	(0.68)	(0.096)	(0.38)	(0.100)	(0.65)	(0.66)	(0.62)
Selection score *	-0.51*	-0.91	-0.31**	0.028	-0.17	-0.32	0.31	0.56
= 1 if passed the selection	(0.28)	(0.83)	(0.14)	(0.46)	(0.15)	(0.81)	(0.77)	(0.79)
Control group mean	0.41	15.8	0.11	3.3	0.11	4.3	3.6	3.7
Ν	252	252	252	252	252	236	235	236
R squared	0.018	0.009	0.016	0.000	0.006	0.007	0.008	0.007

Note: Estimated coefficients are reported. ***, **, and * indicate the 1%, 5%, and 10% levels of statistical significance, respectively. Numbers in parentheses are t-statistics based on heteroscedasticity-robust standard errors.

Table 5. Phist-stage results (Local RDD)	8
--	---

	(1)	(2)	(3)	(4)
	=1 if studied	abroad for 4	Months stu	died abroad
	weeks o	or longer		
OLS	0.40***		5.16***	
(=1 if score is above cutoff)	(0.10)		(1.14)	
IV		0.87***		11.1***
(=1 if passed the selection)		(0.17)		(2.54)
Selection score	0.032	-0.60**	-0.85	-8.96**
	(0.22)	(0.30)	(2.18)	(3.82)
Selection score *	-0.095	0.55*	-1.63	6.58*
= 1 if score is above cutoff	(0.23)	(0.30)	(2.67)	(3.87)
=1 if male	0.022	0.0083	0.11	-0.067
[0-1]	(0.038)	(0.033)	(0.65)	(0.64)
Grade at baseline	-0.021	-0.020*	-0.33	-0.32
[9–21]	(0.014)	(0.012)	(0.20)	(0.21)
=1 if HH income high	-0.10*	0.00050	-3.08***	-1.75**
[0-1]	(0.062)	(0.054)	(0.74)	(0.87)
Baseline English CEFR level	0.014	0.014	0.30	0.29
[1-6]	(0.026)	(0.023)	(0.39)	(0.41)
=1 if no baseline English level	-0.050	-0.028	-0.65	-0.37
	(0.053)	(0.040)	(0.84)	(0.83)
Control group mean	0.	47	3.	55
Ν	2:	52	25	52
R squared	0.308	0.444	0.174	0.137

Note: Estimated coefficients reported. ***, **, and * indicate the 1%, 5%, and 10% levels of statistical significance, respectively. Numbers in parentheses are t-statistics based on heteroscedasticity-robust standard errors. The applied course dummies are included in the regression, but their coefficients are not reported.

Table 6: Mail results (Local RDD)

	(1)	(2)	(3)	(4)	(5)	(6)
	Pooled		1 st follov	v-up only	2 nd follow	v-up only
Panel A: Pragmatic test						
OLS	0.83**		0.60		1.39**	
(=1 if score is above cutoff)	(0.39)		(0.54)		(0.58)	
IV		1.66**		1.12		2.90**
(=1 if passed the selection)		(0.73)		(0.90)		(1.15)
Control group mean	7.	00	6.	93	7.	08
Control group standard deviation	1.	.97	1.	73	2.	18
Ν	3	72	19	94	1	78
R squared	0.211	0.231	0.249	0.278	0.267	0.245
Panel B: International posture						
OLS	0.33**		0.51**		0.16	
(=1 if score is above cutoff)	(0.15)		(0.22)		(0.19)	
IV		0.62**	. ,	0.93**	. ,	0.30
(=1 if passed the selection)		(0.28)		(0.42)		(0.35)
Control group mean	4.	.53	4.	62	4.	66
Control group standard deviation	0.	.90	0.	94	0.	89
N	3	59	19	90	1	69
R squared	0.315	0.302	0.342	0.304	0.324	0.325
Panel C: Perceived competence in L2						
OLS	13.1***		10.9**		14.5***	
(=1 if score is above cutoff)	(3.71)		(5.08)		(5.40)	
ĪV		24.6***		19.8*	× ,	28.0**
(=1 if passed the selection)		(8.45)		(10.6)		(12.1)
Control group mean	5.	3.0	56	5.0	51	1.5
Control group standard deviation	20	0.2	19	9.5	20).6
N	3	59	19	90	10	69
R squared	0.248	0.114	0.320	0.217	0.221	0.079

Note: Estimated coefficients reported. ***, **, and * indicate the 1%, 5%, and 10% levels of statistical significance, respectively. Numbers in parentheses are t-statistics based on heteroscedasticity-robust standard errors. The selection score, its interaction is the treatment dummy, the applied course dummies, and the baseline characteristics, as in Table 6, are included in the regression, but their coefficients are not reported. In columns 1 and 2, the second follow-up dummy is controlled in Panels B and C, the baseline level of the outcome is controlled, but these coefficients are not reported.

Figure 1: Relationship with the running variable



Panel A: Scholarship acquisition (=1 if acquired the scholarship)

Panel B: Stay abroad probability (=1 if stayed abroad for four weeks or longer)



Panel C: Stay abroad duration



Figure 2: The McCray test



Note: The computed *p*-value for the null hypothesis of the absence of bunching near the cutoff is 0.379.

Figure 3: Sensitivity analysis





Panel B: International posture



Panel C: Perceived competence in L2



Note: This sensitivity analysis is based on the pooled data and used OLS estimation, and thus, it is comparable to column 1 in Table 6.