FRIEDRICH-SCHILLER-UNIVERSITÄT JENA

Can there be too much information? Heterogeneous responses to information on benefits from language proficiency

Fabian Könings

Can there be too much information? Heterogeneous responses to information on benefits from language proficiency[*]

Fabian Koenings[†]

Friedrich Schiller University Jena

October 2021

Abstract

Immigrants who have a better command of the host country's language are more likely to be employed and earn higher wages. Using a survey experiment among international students in Germany, I investigate whether information on the monetary benefits of mastering the language of the host country influences the intention to learn that language. The results show heterogeneous responses conditional on the current level of German language proficiency. The intended participation of international students with high German language skills is not affected, students with medium German language skills are positively affected and those with low or no German language skills are negatively affected. For policy makers, seeking to increase the level of language proficiency, this surprising negative effect suggests that there can be too much information.

JEL classification: D01, F22, J24, O15

Keywords: language learning, information experiment, migration, international students

[†] Carl-Zeiss-Str. 3, 07743 Jena, Germany, email: fabian.koenings@uni-jena.de.

1. Introduction

Proficiency in the language of the host country is of key importance for immigrants' social and economic integration (Borjas 1990; Chiswick and Miller 1998; Dustmann and Fabbri 2003; Lochmann et al. 2019). Good language skills make it easier to communicate with local residents and, thus, facilitate social contacts (Bleakley and Chin 2010). Wage benefits of English proficiency have been shown for Australian immigrants (Guven and Islam 2015). Positive effects of being proficient in the host country's language on productivity, employment probability, and wages are also reported for immigrants in the UK (Dustmann and Van Soest 2002) as well as Germany (Dustmann and Fabbri 2003). Furthermore, a more recent study for Germany by Giesecke and Schuss (2019) shows in a simplified cost-benefit analysis that wage gains due to language training (as part of integration courses) outweigh costs in a very short time period (less than three years). For refugees, a particular group of immigrants, the positive effects of language classes on employment and earnings are also supported (Arendt et al. 2020).

Despite the fact that mastering the host country's language has these positive effects, immigrants often do not show sufficiently high levels of proficiency (Isphording 2015). There are several reasons why immigrants' language skills differ. The meta-analysis by Chiswick and Miller (2015) identifies three determinants for learning the language of the host country: exposure, efficiency, and economic incentives. Firstly, language learning is facilitated by increased exposure to the destination language before migration, e.g. if the languages of the two countries are very similar (Isphording and Otten 2013; Raijman 2013), or if a person grew up in a multilingual environment (Budría and Swedberg 2019). Exposure to the language of the host country increases immigrants' language proficiency especially in their early years after arrival (Isphording and Otten 2013). Secondly, people learn a new language with varying efficiency. For example, children or highly educated individuals learn a foreign language more efficiently than adults or less educated persons (Chiswick and Miller 2007; Isphording and Otten 2013). Thirdly, higher economic motivation creates greater incentives to learn a new language. For example, Giesecke and Schuss (2019) show that immigrants in Germany with higher gains in terms of wages and employment are more likely to participate in language training.

To increase social and economic integration, especially in countries where an excess labor demand exists, policy makers might be interested in increasing the language proficiency of immigrants. How can this be achieved? One way, which entails comparatively few costs, is to provide immigrants with information on the economic benefits of learning the host country's language. Given that economic motivation constitutes a key determinant for learning a new language, providing this information might result in a higher intention to learn the language in cases where immigrants do not know about the monetary benefits, they underestimate them, or simply forgot about them.[1] Using an information provision experiment among international students, this study empirically tests this hypothesis.

A sample of international university students was randomly divided into two groups shortly after their arrival in Germany. One group received an information treatment about extra earnings immigrants could realize if they were fluent in German while the other group did not. The study then analyzes how their intention to participate in a German language

---

[1] In cases in which the information is not remembered, the provision makes the information salient and, hence, results in a treatment effect (see e.g.Bleemer and Zafar 2018).

course is affected. The experiment is conducted among international students as this group of immigrants is of particular importance for host country's labor markets in cases of skilled-labor shortage (see e.g. German Federal Ministry for Economic Affairs and Energy 2017). Exogenous potentials from other countries, such as international students, are one possibility to better meet the excess labor demand. But to get employed a good command of the host country's language is of high importance (Geis-Thöne 2019; SVR 2021). Furthermore, literature indicates that insufficient language skills are often a reason why international students have problems to study successfully and do not graduate (see Wisniewski 2018 for a discussion on Germany). An effective information treatment upon arrival in the host country, which motivates international students to increase their language skills, thus, might lead to more international students who graduate and who are then available to fill the gaps on the labor market.

There are several reasons why the information treatment is likely to be effective. First, recent articles employing information provision experiments in varying contexts show that information provision can significantly change preferences (see e.g. Lergetporer et al. 2020; Kuziemko et al. 2015), behavior (see e.g. Chopra et al. 2019) and intentions (see e.g. Bleemer and Zafar 2018). Second, the information is likely to be effective particularly among the group of international students due to a lack of information. Literature indicates that this group misses important information regarding the host country upon arrival (see the qualitative study by Pineda 2018 for Germany). For example, many international students are surprised by the comparatively high cost of living in Germany upon arrival. Third, the monetary and non-monetary costs of language learning are comparatively low for this group because they often receive student discounts and learn languages more efficiently due to their relatively high human capital. These comparatively low barriers for language learning make it more likely that a brief information treatment lets more students decide to participate in a language course.

The results of the present study show that the information provided affects the intention to learn the language heterogeneously. Respondents with a medium level of the German proficiency are (as expected) positively affected by the treatment, while students with a very high level are unaffected. Unexpectedly, individuals with a low level of language competence react negatively to the treatment as they decrease their intention to participate in a language course. In absolute terms, the treatment effects conditional on the level of language proficiency range between -18 to 42 percentage points. These results for international students might be generalizable to other groups of immigrants, e.g. those with the primary purpose of working in the host country. The provided information on the benefits of language proficiency on the labor market might be even more important for these immigrants and outweigh the higher costs of language learning for non-students.

These findings are important for policy makers with the intention to increase the language competence of immigrants via information campaigns. They should take into account the above mentioned encouraging and discouraging effects. Especially the discouraging effect on individuals with a low level of proficiency in the host country's language should be heeded as they might be the most important target group for this policy intervention. Thus, information campaigns have to be either targeted towards groups with specific skill levels or the information has to be adjusted according to the skill level to possibly result in exclusively higher levels of language proficiency.

The structure of this paper is as follows: In the next section, I present the survey, the design of the experiment as well as the empirical strategy. The third section discusses the randomization in the experiment and presents the findings on the treatment effect across all groups. In the fourth section, I investigate heterogeneous treatment effects by the current level of host country's language proficiency. The fifth section discusses the findings and the last section concludes.

2. Data and Empirical Strategy

2.1. The Survey

To test whether information on monetary benefits from language proficiency on the labor market increases the intention to participate in a language course, international students at a medium-sized university in Germany were surveyed. The fully-fledged university under study is representative for an analysis of international students in Germany with respect to its ratio of international students to all students (13.84%). This share is close to the average share of foreign students at German universities, which is 13.44%.[2] (For a further discussion on the representativity of the resulting sample see subsection 2.4.)

The survey gathered data on the intention to participate in a German language course and further characteristics for 324 international students.[3] According to Ajzen (1991)'s theory of planned behavior, intention is almost always a necessary precondition for an actual behavior. Furthermore, the intention to choose a particular education has been shown to be a very good predictor of future education choice (see e.g. Jacob and Linkow 2011). Therefore, I employ the intention to participate in a language course as a meaningful proxy for actual language course participation. Besides this question on the intended participation in a German course, the survey collects information on sociodemographic characteristics, information on study programs, expected problems in Germany, students' plans for the future, as well as further individual characteristics concerning the German language such as the current knowledge of German.[4]

To get a first visual impression of the data and to provide an exemplary distribution of German language proficiency of non-Germans, I plot the distribution of the respondents' self-assessed German proficiency in Figure 1. The level of German is measured on a six-point scale ranging from 0 "no knowledge" to 5 "very good knowledge". International students with German as their mother tongue, e.g. Austrians, are excluded from the data because they are not expected to participate in a German language course. The figure reveals several insights. First of all, there is a considerable proportion of respondents who have a low to non-existent level of German language proficiency, which shows the necessity to participate in (additional) German courses. Second, there are also some respondents who already have a relatively high level of German proficiency, resulting in a very heterogeneous distribution of language proficiency. Third, since the respondents were free to choose whether they answer the questionnaire in German or in English, Figure 1

---

[2] These figures are own calculations based on data from the German Federal Statistical Office 2019.

[3] Overall, data on the intention to participate in a German language course for 415 international students was collected. Due to item-non-responses I focus in the analyses on 324 individuals. T-tests comparisons for pre-treatment variables between available out-of-sample and in-sample observations show no significant systematical difference.

[4] The questionnaire is available upon request.

additionally shows the distribution of the language of the questionnaire. Choosing the German over the English questionnaire highly correlates with the self-assessed measure of German proficiency, indicating a certain meaningfulness for the more detailed self-assessed measure of language proficiency. Finally, even among the respondents with a very high level of German knowledge (i.e. a level of 4 or 5) there are some who chose the English over the German questionnaire. Hence, the sample of English questionnaires is spread along all dimensions of German language proficiency. In contrast, the German questionnaire sample is almost exclusively concentrated on very good German knowledge.
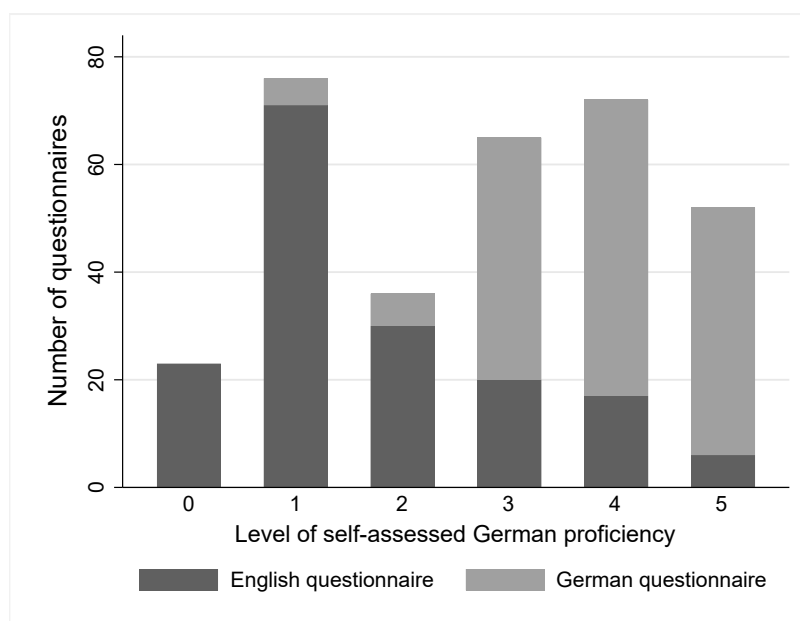

Figure 1: Distribution of self-assessed German proficiency

Notes: The self-assessed German proficiency is measured on a 6-point scale ranging from "no knowledge" (0) to "very good knowledge" (5).

The survey took place from 24 September 2018 to 5 October 2018 which constitutes the beginning of the winter semester 2018. The investigation period coincides with the early weeks after students' arrival in Germany, thus, the foreign students still have limited information and experience on the German labor market. To maximize the response rate, students were surveyed by means of a paper-and-pencil questionnaire. Overall, the response rate among all newly enrolled international students was 63%.

2.2. The Experiment

To investigate how information affects international students' intention to participate in a language course, the survey includes an experiment. All questionnaires contain the question of main interest for this study, i.e. whether the international students plan to participate in a German language course during their studies. Before this question, the respondents were confronted with a sentence that varied between two types of questionnaires. The two different types of questionnaires were randomly distributed such that one half of the respondents – the treatment group – received a sentence with information about the monetary benefits of proficiency in German for lifetime earnings in Germany. The information treatment reads as follows:

"Economic research shows that German language proficiency increases lifetime earnings of immigrant workers in Germany by 5% to 15%, corresponding to 47,000 to 142,000 Euro."[5]

The other half of respondents – the control group – received a questionnaire with a 'placebo-sentence' containing no information about the potential gains from proficiency in the host country's language:

"In the following, we would like to learn more about your experiences prior to coming to Germany. Furthermore, we are interested in your plans for the near future."

The two different types of questionnaires present the sentence containing the information and the 'placebo-sentence' in an identical format. Similar to Card et al. (2012), I chose a placebo instead of no additional sentence in order to minimize the chance of confounding factors influencing the treatment effect. Students could, for example, generally react to an additional sentence (without a question) within a questionnaire. The appearance of a sentence in the treatment and control questionnaire ensures that the treatment effect is not influenced by such general reactions.

The methodology of information provision experiments has been applied to a wide variety of topics in order to identify causal effects of information provision (see e.g., Cruces et al. 2013, Elias et al. 2015 or Kuziemko et al. 2015). Some studies that conduct information provision experiments investigate beliefs about the provided treatment before the information is revealed (see e.g., the investigation of pre-treatment beliefs by Chopra et al. 2019). Having information on prior beliefs provides the possibility of analyzing the updating process of beliefs through the information given. While in the present study data on pre-treatment beliefs of the monetary gains from language proficiency on the labor market would be very insightful, a paper-and-pencil survey does not allow for such an investigation. It cannot be credibly asserted that respondents do not include the information treatment in their pre-treatment beliefs.

2.3. Econometric Model

Since randomization between treatment and control group ensures a causal interpretation of the effect of the treatment on the intention to participate in a language course, the estimation model for individual $i = 1, ..., N$ assumes the form

$$int_i{}^* = \alpha + \beta \, Treatment_i + \gamma' X_i + \epsilon_i \tag{1}$$

where $int_i{}^*$ is the latent propensity to participate for individual i, which is based on the binary intention to participate int_i such that $int_i^* > 0$ if $int_i = 1$ and $int_i^* \leq 0$ if $int_i = 0$ holds. $Treatment_i$ indicates whether individual i receives the treatment. The vector X_i contains individual specific control variables (see the next subsection for a detailed description of the variables included) and ϵ_i is the idiosyncratic error term. I use a standard probit model framework, which assumes a cumulative normal distribution function to estimate the latent propensity to participate in a language course.[6]

---

[5] Percentage values are taken from Dustmann and Van Soest (2002). Absolute values are based on own calculations referring to the percentage values by Dustmann and Van Soest (2002) and average lifetime earnings estimations for university graduates in Germany by Piopiunik et al. (2017).

[6] All analyses in this paper that use a probit model are also conducted using a linear probability model (LPM) or logit model. For the results using the LPM or the logit model, see Appendix B.

Due to the randomization, the simple difference between the average propensity to participate of the treatment group and the control group already provides the causal impact of the treatment. Including additional control variables, X_i, may generate a higher precision of the causal impact (Angrist and Pischke 2009). If the treatment is uncorrelated with the control variables, the estimated coefficient β changes only marginally between models with and without control variables. For traceability, I decided to always report both regression results - with and without control variables.

2.4. Control Variables and Descriptive Statistics

The control vector X_i includes several variables which may be associated with international students' intention to participate in a German language course.[7] Next to demographic characteristics such as age, gender, and the region of origin, it contains individual characteristics such as having a partner, being risk averse, and being patient. Furthermore, the current self-assessed level of German proficiency is part of the vector as individuals with a high level of German might be less inclined to participate in further language courses. In addition, an indicator on whether the students expect linguistic barriers during their studies is included. Since the desire for social interaction with locals is an incentive to learn the host country's language, an indicator on whether the student values social contacts as important is a further control variable (see e.g. Bleakley and Chin (2010) for the importance of language proficiency for social contacts in the host country).

Moreover, study-related variables enter the control vector X_i. Information on whether a German course is recommended for the study program, whether one is enrolled for a relatively short period of time (Erasmus student or similar), and the type of degree (Bachelor or Master) are part of the vector. Since Chiswick and Miller (2006) point out the importance of the time period that immigrants want to stay in the host country for the intention to learn the host country's language, it is likely that international students who plan to stay in Germany are also more inclined to learn the German language. Therefore, the intention to stay in Germany after graduation weighted by the certainty of this decision is another control variable.[8]

In the upper part of Table A1 of Appendix A, I display, among others, average values for the control variables to provide some descriptive statistics on the data used. The first column shows the mean values across all respondents. The average age in the sample is 23 and there are more female than male respondents. The main region of origin is Asia followed by Europe. The majority of students has no partner, considers themselves risk neutral to risk loving and patient. As you can already see in Figure 1, the average German language proficiency lies between basic and good with a value close to 3. Linguistic barriers during the studies are expected by more than half of the students. Almost all respondents consider social contacts as important. Around 40% of the students are part of the Erasmus or a similar program and slightly less than 50% are enrolled in a Master program. With varying degrees of certainty, about half of the international students plan to stay in Germany after their studies.

To assess the representativity of the sample, in Table A2 of Appendix A I compare the

---

[7] The scales on which the control variables are measured are listed in the notes to Table A1 of Appendix A.
[8] For a more detailed analysis on the intention to stay of international students and the uncertainty involved, see Koenings et al. (2021).

distribution across gender, age, region of origin, the type of degree and the field of study in the sample to the distribution of these variables for all newly enrolled international students in the winter semester 2018 at the university under study (column (2)) and all German universities (column (3)). The comparison of the sample with column (2) shows that the sample is fairly representative for the university under study. Except for the sample being on average slightly younger, having a slightly lower share of students in a Master program, and a slightly larger share of students studying 'Humanities', the numbers for the sample and the university are similar. Therefore, contrasting the sample to the whole universe of international university students in Germany reveals some particularities of the university under study. The university under study has relatively less female students, less European and more Asian students, as well as more students in a Master program. Furthermore, students of 'Mathematics, Natural Sciences' are overrepresented while students studying 'Engineering' are underrepresented.

3. The Treatment Effect

This section analyses the effect of the treatment and is split into two parts. First, I test whether the random provision of the two different types of questionnaires arbitrarily sorted the respondents into treatment and control group. Verification of the randomization allows for causal interpretation of the treatment effect. Second, results on the average treatment effect are displayed and discussed.

3.1. Verification of Randomization

The questionnaires were arranged before the start of the survey such that the treatment version and the control version always alternated. Of 324 respondents, 49.7% received the treatment and, hence, constitute the treatment group.

To assure that the groups to be analyzed indeed resulted in randomized groups, I test on differences in group means among the included control and further pre-treatment variables. Table A1 in Appendix A shows that group means of the treatment and the control group do not differ statistically for 42 out of 48 variables. The few exceptions are the following: The treatment group is slightly younger and more risk averse than the control group. Also, it has slightly more students being part of an Erasmus or a similar program and less students being enrolled in a Master program. Slightly more respondents in the control group agree that learning German is helpful for social integration and makes life easier in Germany.

The large fraction of statistically insignificant differences verifies the random provision of the information treatment. This implies that the effect of the treatment on the intention to participate in a language course can be interpreted as causal.[9]

3.2. Results

Figure 2 plots the average intention to participate in a language course for the control and treatment group separately. As hypothesized, the intention to participate is higher among

---

[9] The few statistically significant differences between treatment and control group characteristics are another reason why all regression analyses are also conducted including control variables.

the treatment group with 71.4% compared to the control group with 68.7%. However, the intended participation in both groups is high and the difference between the groups is relatively small. The 90% confidence intervals (CIs) in Figure 2 indicate that this difference is not statistically significant. This result is mirrored by the positive but insignificant coefficient for the treatment in column (1) of Table 1 which provides the estimation results of Equation (1) without control variables. When controlling for additional variables, the coefficient remains insignificant (column (2) of Table 1). These results show that the provided information does not have an effect on the intention to participate when the respondents are examined all together.
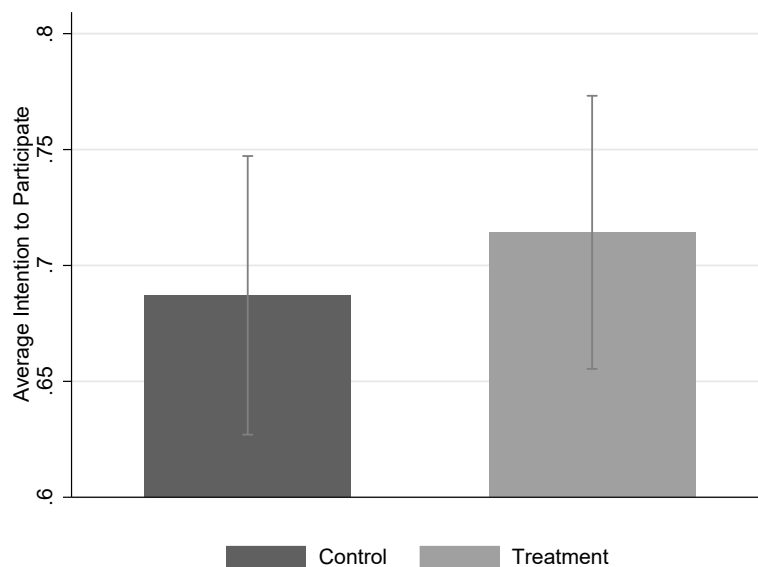
Figure 2: Treatment effect

Notes: The average intention to participate in a German language course for the treatment and the control group is displayed. The figure includes 90% CIs.

The coefficients for the control variables show that respondents who are more patient, have a higher intention to participate in a German course. The level of self-assessed German language proficiency is significantly negatively associated with the intention to participate. As expected, students with an already high proficiency of the German language are less likely to enroll in an additional German course. Respondents, who expect less linguistic barriers during their studies are, however, more inclined to participate. Students enrolled in a Bachelor degree have a higher intention to participate than students enrolled in other degrees.

4. Heterogeneous Treatment Effects by the Level of German

Several studies have shown that subgroups may react heterogeneously to an information treatment (see, for example, the studies by Card et al. 2012 or Cullen and Perez-Truglia 2018). These studies find that the effect of the treatment exists only for particular subgroups of the treated while it does not for others. Cullen and Perez-Truglia (2018), for example, find that information about the high salaries that managers earn makes employees work harder, if they are likely to become managers in the future. However, the

Table 1: Treatment effect with and without control variables

	(1)		(2) with control variables	
Treatment	0.078	(0.147)	0.068	(0.180)
Control variables				
Age			−0.024	(0.030)
Female			−0.035	(0.197)
Region				
Americas			0.326	(0.685)
Asia			−0.217	(0.581)
Europe			−0.122	(0.593)
Having a partner			−0.248	(0.262)
Risk aversion			−0.044	(0.031)
Patience			0.099***	(0.032)
Self-assessed German proficiency			−0.334***	(0.081)
Social contact is important			0.127	(0.110)
Linguistic barriers			−0.230***	(0.079)
German course recommended			0.434	(0.325)
Erasmus or similar			0.326	(0.217)
Bachelor program			0.469**	(0.237)
Master program			0.356	(0.238)
Staying with certainty			0.023	(0.037)
Constant	0.488***	(0.103)	1.440	(1.034)
Observations		324		324
Pseudo-R^2		0.001		0.300

The dependent variable is the intended participation in a German course. The reference category for 'Region' is Africa. Standard errors are displayed in parentheses. The stars indicate the following: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

treatment has no effect on employees that do not have a reasonable chance of becoming managers soon. Furthermore, Haaland et al. (2021) discuss the possibility that an average treatment effect of a survey experiment across all participants may be muted when the information results in opposite treatment effects conditional on prior beliefs.

In the present study, it is likely that the provision of the information has heterogeneous effects on students depending on their current level of German language proficiency. If, for example, the level of German is already very high, then the treatment is more likely to be ineffective as participation in a further German course is not necessary to earn the potentially higher salaries mentioned in the information. If this is the case, high levels of German language proficiency might cause the insignificant treatment effect seen in the previous section.

Therefore, in the following subsections, I examine in detail possible heterogeneous treatment effects conditional on the level of German language proficiency.

4.1. Empirical Strategy

To test for heterogeneous treatment effects by the level of German language proficiency, I run the following estimation equation

$$int_i^* = \alpha + \beta_1 \text{Treatment}_i + \beta_2 \text{Proficiency}_i + \beta_3 \text{Proficiency}_i * \text{Treatment}_i + \gamma' X_i + \eta_i \quad (2)$$

where $Proficiency_i$ is the self-assessed level of German language proficiency of individual i measured on a 6-point scale as shown in Figure 1.[10] X_i contains the same set of control variables as in Equation (1). η_i is the idiosyncratic error term.

Again, I use a standard probit model framework in order to estimate the latent propensity to participate int_i^*. The signs of the coefficients β_1 and β_2 now provide the general direction (positive or negative) on the intention to participate for being treated and for higher German proficiency, respectively. The sign of β_3 provides the additional information on how the treatment effect changes when the level of German proficiency increases. If, as hypothesized above, the treatment has a stronger effect on students with lower levels of German proficiency, β_1 will be positive and β_3 negative, showing a smaller treatment effect for higher levels of German proficiency.

4.2. Heterogeneous Treatment Effects

Columns (1) and (2) of Table 2 show the estimation results for Equation (2) with and without control variables. Surprisingly, the treatment itself exerts a negative (almost) statistically significant impact, while the interaction between the level of German language proficiency and the treatment has a positive statistically significant coefficient. This means that the provided information has an unintentional negative effect on students with a low

---

[10] Throughout the estimations I treat the self-assessed German language proficiency as a continuous variable. This is due to comparative reasons as in subgroup analyses the number of observations per language level can be relatively small, resulting in omitted coefficients. To take the particular level of German proficiency into account, I visualize treatment effects per language level in the last step of the analyses. Furthermore, I conduct verification analyses employing discrete variables, e.g. choosing the German over the English questionnaire, as proxies for the German language proficiency.

level of German proficiency. The positive coefficient for the interaction term shows that this negative impact decreases with higher levels of German proficiency.[11]

To verify this unexpected result, I run the same analysis as presented in columns (1) and (2) of Table 2 but use another indicator for the level of German proficiency. Figure 1 in subsection 2.1 shows that the level of German proficiency highly correlates with choosing the German over the English questionnaire. Thus, for this verification check I substitute the level of German language proficiency by having chosen the German (N=157) over the English (N=167) questionnaire. Columns (3) and (4) of Table 2 display this verification check with and without control variables. As the coefficient for the interaction term as well as the coefficient for the treatment are far from being statistically significant, this analysis cannot confirm the finding that the treatment exerts a negative effect on students with a low level of German proficiency which decreases with higher language proficiency.[12]

Why are there heterogeneous treatment effects by the level of German language proficiency but not for having completed the German instead of the English questionnaire?

Another look at Figure 1 in subsection 2.1 shows that the joint distribution of the self-assessed German proficiency level and choosing the German over the English questionnaire might be used to categorize the students into different groups of 'actual' German language proficiency. One group is formed by students who chose the German over the English questionnaire. In addition to choosing the German over the English questionnaire, which demands already a good command of German, these students rate themselves on average with a high level of self-assessed German language proficiency. Thus, everyone who chose the German questionnaire is considered to have a *high* level of German. The students that chose the English over the German questionnaire, in turn, exhibit on average lower self-assessed levels of German proficiency. However, these levels are more heterogeneously distributed compared to the German-questionnaire sample. Hence, students that chose the English over the German questionnaire can be categorized into two groups of 'actual' German language proficiency: a group with *medium* German skills (i.e., the ones choosing the English questionnaire but indicating a high level of self-assessed language proficiency) and a group with *low* German skills (i.e. the ones who chose the English questionnaire and indicate a low level of self-assessed language proficiency).

Now, returning to the hypothesis that the information treatment is likely to be ineffective for students with an already high level of German proficiency (i.e. students that chose the German over the English questionnaire) I run Equation (2) separately for German and English questionnaires. Table 3 displays the corresponding estimation results. Columns (1) and (2) show the estimated coefficients for German questionnaires only. As expected

---

[11] From a theoretical point, it is possible that the treatment might also act heterogeneously for students who want to stay in Germany and those who do not. Furthermore, students from countries whose official languages are more similar to German, i.e. have a lower linguistic distance, might be more affected by the treatment, as it is easier for them to learn German. However, the replication of the differential analyses for the intention to stay in Germany and linguistic distance does not provide evidence for these lines of argument. (For the analysis on linguistic distance the indices by Adserà and Pytliková (2015) are used.)

[12] Employing a logit model or a LPM results in the same findings for columns (3) and (4) of Table 2. The treatment effects by the level of German proficiency in columns (1) and (2) are slightly less statistically significant when a logit model or a LPM is employed. These models suggest that the heterogeneous treatment effects measured when the whole sample of respondents is employed might be not that strong. For the corresponding estimation results using the logit model or LPM see B3 and B4 in Appendix B, respectively.

Table 2: Heterogeneous treatment effects by levels of self-assessed German proficiency and German questionnaire

	(1) Proficiency	(2) Proficiency with control variables	(3) German questionnaire	(4) German questionnaire with control variables
Treatment	-0.885^*	-0.766	-0.152	0.029
	(0.452)	(0.471)	(0.257)	(0.287)
German proficiency	-0.653^{***}	-0.489^{***}		
	(0.104)	(0.119)		
Treatment $\times$ German proficiency	0.288^{**}	0.260^*		
	(0.126)	(0.134)		
German questionnaire			-1.350^{***}	-0.680^{**}
			(0.236)	(0.290)
Treatment $\times$ German questionnaire			0.297	0.075
			(0.326)	(0.364)
Control variables		Yes		Yes
Observations	324	324	324	324
Pseudo-R^2	0.221	0.310	0.153	0.276

The dependent variable is the intended participation in a German course. The control variables include age, gender, region of origin, an indicator for having a partner, and the level of risk aversion, and of patience. They further include indications on whether linguistic barriers are expected, whether social contact is important, whether a German course is recommended for the study program, being part of Erasmus or a similar program, being enrolled in a Bachelor or Master program, and the intention to stay in Germany after graduation multiplied by the level of certainty. Standard errors are displayed in parentheses. The stars indicate the following: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

for this subgroup, the treatment does not have a significant impact. Furthermore, there are no heterogeneous effects by the level of self-assessed German proficiency. The results for the English-questionnaire sample with and without control variables are displayed in columns (3) and (4) of Table 3. The treatment as well as the interaction term turn out statistically significant, showing that the heterogeneous treatment effects by the level of German (observed in Table 2) are rooted in the English-questionnaire sample.[13]

4.3. Robustness Checks

To verify the results for the English-questionnaire sample, I conduct two robustness checks. First, I substitute the control group with a group of international students who were not subject to any experiment. The data on the new control group stems from the same survey conducted one year earlier among incoming international students without the experiment and, hence, without any information treatment and no 'placebo-sentence'. Substituting the old control group (with 'placebo-sentence') with the new one (without any additional sentence) reveals whether the findings are driven by how the 'placebo-sentence' is framed.[14]

In a second robustness check, I use being enrolled in a German-speaking study program instead of an English-speaking one as a proxy for German language proficiency. While it can be argued that the self-assessed level of German language proficiency might not reflect the real level of German proficiency, being enrolled in a German-speaking program is a more reliable indicator for German proficiency. This is due to the fact that all students who do not have a German university entrance certificate nor at least a 3 years bachelor degree in Germany must submit a document verifying an appropriate level of their German ("Deutsche Sprachprüfung für den Hochschulzugang" (level 2) or equivalent).[15]

The results of the first robustness check using a new control group are shown in column (5) and (6) of Table 3. Columns (7) and (8) of the same table display the results for the second robustness check. They both confirm the results of a negative treatment effect on students with a very low level of German proficiency while this negative effect decreases with the level of German proficiency.[16]

---

[13] To interpret the effect of the treatment as causal within the sample of English questionnaires the randomization between treatment and control group for this sample has to be verified. The control group consist of 81 and the treatment group of 86 respondents. The randomization can still be verified as out of 48 variables only 3 differ significantly between treatment and control group (see Table A3 of the Appendix A). The control group does not contain an international student who learned German within the family. In addition, the share of students enrolled in the field of humanities and considering learning German to be creating more opportunities in the future is slightly smaller in the treatment group than in the control group.

[14] To interpret the results using the new control group as causal, again, the randomization between the new control group and the treatment group has to be verified. I compare group means for students who choose the English over the German questionnaire in Table A4 of Appendix A. In 35 of 49 cases, there are no statistically significant differences between the two groups. Important differences are controlled for via the control variables included in the estimations.

[15] The indicator on being enrolled in a German-speaking study program is based on the question "What is the main language of the study program?" and cross-checked with an overview on all German-speaking study programs (https://www.uni-jena.de/Studienangebot, accessed on 09.08.2021).

[16] When employing a logit model or a LPM results remain largely the same (see Tables B5 and B6 in Appendix B).

Table 3: Heterogeneous treatment effects by self-assessed German proficiency: Different samples and robustness checks

	(1) German questionnaires: Proficiency	(2) German questionnaires: Proficiency with control. variables	(3) English questionnaires: Proficiency	(4) English questionnaires: Proficiency with control variables	(5) English questionnaires: New control group group	(6) English questionnaires: New control group with control variables	(7) English questionnaires: German program	(8) English questionnaires: German program with control variables
Treatment	0.429	1.548	−1.305**	−1.758**	−1.032**	−1.803***	−0.637*	−0.562
	(0.940)	(1.178)	(0.567)	(0.793)	(0.453)	(0.589)	(0.377)	(0.468)
German prof.	−0.540***	−0.434**	−0.562***	−0.641***	−0.436***	−0.636***		
	(0.174)	(0.214)	(0.170)	(0.225)	(0.120)	(0.168)		
Treatment×German prof.	−0.059	−0.334	0.498**	0.833***	0.346**	0.679***		
	(0.235)	(0.287)	(0.206)	(0.300)	(0.171)	(0.229)		
German progr.							−0.986**	−0.989*
							(0.415)	(0.588)
Treatment×German progr.							0.970*	1.585**
							(0.568)	(0.732)
Control variables		Yes		Yes		Yes		Yes
Observations	157	157	167	167	201	201	167	167
Pseudo-R^2	0.126	0.261	0.121	0.341	0.115	0.261	0.054	0.284

The first two columns use the German-questionnaire sample and the succeeding columns use the English-questionnaire sample. The dependent variable is the intended participation in a German course. The control variables include age, gender, region of origin, an indicator for having a partner, and the level of risk aversion, and of patience. They further include indications on whether linguistic barriers are expected, whether social contact is important, whether a German course is recommended for the study program, being part of Erasmus or a similar program, being enrolled in a Bachelor or Master program, and the intention to stay in Germany after graduation multiplied by the level of certainty. Columns (4) and (8) do not control for the region of origin as the seven African students, who all want participate in a German course, would be dropped from the sample and, hence, the numbers of observations would not align with columns (3) and (7), respectively. However, controlling for the region of origin also in these columns does not affect the results. Standard errors are displayed in parentheses. The stars indicate the following: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

4.4. Effect Sizes

To visualize the magnitude of the heterogeneity, I first calculate the predicted propensities of having the intention to participate in a German course for the different levels of German, with and without treatment. I do this separately for the German- and the English-questionnaire sample.[17] (For the results see Figure A1 of Appendix A.) Second, I take the difference of the predicted propensities between treatment and control group per level of German and plot the resulting treatment effects in Figure 3.

The left-hand side of Figure 3 shows the statistically insignificant treatment effects in the German-questionnaire sample, i.e. the ineffective information provision among respondents with an actual *high* level of German. The right-hand side of Figure 3 shows the heterogeneous treatment effects which can be observed in the English-questionnaire sample. It shows that the treatment exerts a negative impact on the subgroup with a *low* level of actual German proficiency (i.e. English questionnaire and a self-assessed language level of 0 or 1). This negative impact diminishes with increasing self-assessed German levels and turns significantly positive for the students that fall into the category of *medium* level of German knowledge (i.e. English questionnaire and a self-assessed language level of 4 or 5). The effect sizes range between -18 to 42 percentage points.

Figure 3 also shows why a treatment effect across all levels of German language proficiency in the previous section was not identified. The effects into opposite directions cancel each other out resulting in a so-called muted average treatment effect.

5. Discussion

The results show that the information on wage benefits of German language proficiency exerts heterogeneous effects on the intention to participate in a German course depending on the current 'actual' level of German. The treatment does not have an effect on students with an actual high level of proficiency (i.e., students who choose the German over the English questionnaire). The provision of information has a positive effect on students with a medium actual level of German proficiency (i.e., the ones choosing the English questionnaire and indicating a high level of self-assessed German knowledge). Unexpectedly, the treatment has a negative effect on students with an actual level of German proficiency that is low, i.e. students who chose the English questionnaire and self-assess their German level to be low or non-existent.

The reason for these heterogeneous effects is probably a cost-benefit analysis with different results depending on the students' current level of German language proficiency. The observed negative effect on intended course participation, for example, likely lies in a combination of high costs to learn a new language from scratch and low expected returns given the costs. To become proficient in a language (as required in the provided information) involves on average higher costs for those who only have low or no proficiency in the language compared to those who already have some (medium) proficiency. Taking these higher costs into account, the expected returns (given in the treatment) might seem low, and thus, students – after receiving the information – decide to allocate their time to other activities. As a consequence, a discouraging effect of the information on students with a

---

[17] The calculations are based on the estimation results including control variables, i.e., column (2) and column (4) of Table 3.

Figure 3: Treatment effect sizes by self-assessed level of German proficiency

Notes: The left-hand side displays the treatment effects for each level of self-assessed German proficiency in the German-questionnaire sample. The right-hand side displays the treatment effects for each level of self-assessed German proficiency in the English-questionnaire sample. For the German-questionnaire sample there is no value for a self-assessed German proficiency level of 0 since there are no observations in this category. Calculations are based on the estimations shown in columns (2) and (4) of Table 3. The plots are displayed with 90% CIs.

low level of German proficiency can be observed.

In contrast, for students who already have at least a medium level of German proficiency (i.e., students who chose the English over the German questionnaire but assess their level of German language proficiency higher than the students with no or only low knowledge) the treatment exerts a positive effect. This might be because costs to become fully proficient are lower for students with a medium level than with a low level. In this case, the cost-benefit analysis encourages students to participate in a language course. Students, who already have a high level of German language proficiency are neither encouraged nor discouraged by the information provided as they already have a good command of the German language. Hence, they might not feel the need to further invest into a language course.

Some might argue that the observed heterogeneous effects, and in particular the negative effect, might be due to a low level of understanding of the rather complex treatment, which involves absolute and relative numbers, and/ or misunderstanding of the question on the intended course participation. Hence, the heterogeneous results might rather show a cognitive ability effect. However, three factors go against this explanation. First, a misunderstanding of the treatment and question on intended course participation most likely would not result in statistical significant effects in a positive or negative direction. One would rather expect a fuzzy insignificant effect. Second, a linguistic misunderstanding is unlikely as well. The respondents could choose the language of the questionnaire (i.e., German or English). Furthermore, they are students at an university where lectures are given in either German or English. Therefore, it can be assumed that they all have a sufficient command of one of these languages to follow lectures and, thus, understand the questionnaire. Third, the level of understanding of the treatment might be correlated with the level of cognitive ability and, thus, causing the heterogeneous results observed. However, using the self-assessed level of English proficiency (excluding native English speakers) as a proxy for cognitive ability, does not present evidence for this explanation. The analysis on differential effects for different levels of self-assessed English proficiency does not show statistically significant (heterogeneous) results (see the corresponding estimations in Table A5 of the Appendix A).

To derive policy implications it is always important to discuss the underlying mechanisms of the results observed. Next to the discussion on why heterogeneous effects are observed, it can further be discussed why the information provision in general exerts statistically significant effects. In this context some scholars identify the total lack of information or the updating of existing information as the driver of the effective information provision (e.g. Cruces et al. 2013 or Lergetporer et al. 2020). Others identify salience, i.e. putting a focus on a particular topic via the information, as the cause of an information provision effect (see e.g. the discussion in Cacciatore et al. 2016 or Bleemer and Zafar 2018). As explained in subsection 2.2, due to the paper-and-pencil nature of the survey, no prior beliefs on the information provided can be investigated and, thus, no empirical analyses can be conducted here. However, as already pointed out in section 1, Pineda (2018) shows that international students lack certain information regarding Germany upon arrival, making the mechanism of belief-updating more likely for the present case. However, future research should empirically evaluate this as it affects the optimal timing of the information provision (Bleemer and Zafar 2018).

I argued that the information provision is likely to be effective for international students because of their lack of important information regarding Germany and their comparatively low monetary and non-monetary costs to learn a language. Now, the question can be raised to which extent the present analysis and its results can be generalized for other groups of immigrants. While a priori (without another group of immigrants having been subject to the same experiment) any statements about the external validity of the experimental results are speculative, from a theoretical point of view it is ambiguous if the results change for other groups of immigrants. Uebelmesser and Weingarten (2017) state that students probably have more motivation to learn German than other groups of immigrants, but it is unclear if this results in larger information treatment effects on students. The lower costs of learning the language for international students (due to low-priced course fees and higher efficiency in learning) very likely increase the effect (in cases of low current language proficiency: diminishes the discouraging effect) of the small informational intervention as it reduces the barriers to participate in a language course. Other groups of immigrants are confronted with higher costs of learning the language, which, in turn, may result in smaller (in cases of low current language proficiency: more negative) information provision effects. However, for immigrants that come to Germany with the primary purpose to work and earn a living, the information on monetary benefits from language proficiency might be of higher importance which may result in larger information provision effects (in cases of low current language proficiency: smaller discouraging effects). With respect to the argument of international students' lack of important information, other groups of immigrants newly arriving in Germany might also have limited information about the country. This is especially likely when the preparation time before migration is short such as it is often the case for refugees. In sum, the results of the experiment for other immigrant groups are difficult to predict and, therefore, are an interesting topic for future investigations.

Treatment effects in information provision experiments might be caused by experimenter demand effects, which means in the present case that respondents would answer positively to the question on the intention to participate in a language course after receiving the information. They would think they are expected to answer the question positively as this is what the experimenter wants (see for a discussion on experimenter demand effects Haaland et al. 2021). However, it is unlikely that the present findings of heterogeneous treatment effects are due to experimenter demand. If experimenter demand were the reason for the results, I would expect positive treatment effects only.

6. Conclusion

Literature shows that immigrants' proficiency in the host country's language is very likely to have positive effects on their social and economic integration (Borjas 1990; Chiswick and Miller 1998; Dustmann and Fabbri 2003). However, heterogeneous levels of the host country's language proficiency of immigrants are observed. Therefore, this study investigates one relatively inexpensive way to incentivize more immigrants to learn the host country's language. To do so, I conduct a randomized survey experiment among newly arrived international students in Germany and test whether information on monetary benefits from language proficiency increases the intention to participate in a language course.

The results show that the treatment effect is heterogeneous across groups. The provided information does not have an effect on the intended participation in a language course of students who already have a good command of the German language. The treatment positively affects students with an intermediate level of the host country's language, i.e. students who choose the English over the German questionnaire but assess their German knowledge to have at least some basic level. Unexpectedly, the treatment has a negative effect on the intention to participate in a language course for individuals with low language proficiency. As the treatment effects work in opposite directions, the effect over all German levels is muted. This result on heterogeneous effects is potentially due to a cost-benefit analysis which leads to a discouraging effect for students with a low level of German proficiency as these students are confronted with high costs to become proficient (as required in the information provided to be rewarded with the monetary benefit).

This study focuses on international students due to their particular importance for host countries' labor markets confronted with skilled-labor shortage if the students have a good command of the host country's language. The results might be international student specific due to their identified lack of important information on the host country upon arrival (Pineda 2018) and comparatively low costs for learning a language. However, it is unknown whether the results change for other groups of immigrants as, for example, immigrants coming with the purpose to work in the destination country might value the information as more important outweighing higher costs for language learning. This study cannot empirically explain whether the treatment effects observed are due to belief-updating or salience while the former is more likely to be the case.

My results are of interest to policy makers as they show, on the one hand, that a relatively small and low-cost information provision can increase the intention to participate in a language course. On the other hand, and maybe even more importantly, my results show that there can also be too much information in the sense that the information provision results in a discouraging instead of the intended encouraging effect. Specifically, the analysis shows that campaigns on the monetary benefits of being proficient in the host country's language might result in unforeseen negative effects for those that might be of most interest, i.e. those with a very low level of the host country's language. When policy makers want to conduct such kind of interventions to increase the level of host country's language proficiency, they have to take this possible negative effect into account and adjust their intervention targeting accordingly. In addition, from a cost-benefit point of view it might be worthwhile to ration the provision of the information further and leave out those on whom the intervention is likely to be not effective, i.e. individuals who already have a good command of the German language. To design appropriate information campaigns provision further research is necessary to show whether the results are generalizable to other groups of immigrants and whether the treatment effect is due to salience or belief-updating.

References

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

Ajzen, Icek (1991). "The theory of planned behavior". *Organizational Behavior and Human Decision Processes 50*(2), 179–211.

Angrist, Joshua D. and Jörn-Steffen Pischke (2009). *Mostly harmless econometrics*. Princeton, New Jersey: Princeton University Press.

Arendt, Jacob Nielsen, Iben Bolvig, Mette Foged, Linea Hasager, and Giovanni Peri (2020). "Integrating refugees: language training or work-first incentives?" *NBER Working Paper 26834*, National Bureau of Economic Research, Cambridge (Massachusetts).

Bleakley, Hoyt and Aimee Chin (2010). "Age at arrival, english proficiency, and social assimilation among US immigrants". *American Economic Journal: Applied Economics 2*(1), 165–192.

Bleemer, Zachary and Basit Zafar (2018). "Intended college attendance: evidence from an experiment on college returns and costs". *Journal of Public Economics 157*, 184–211.

Borjas, George J. (1990). *Friends or strangers: the impact of immigrants on the U.S. economy*. New York: Basic Books.

Budría, Santiago and Pablo Swedberg (2019). "The impact of multilingualism on host language acquisition". *Empirica 46*(4), 741–766.

Cacciatore, Michael A., Dietram A. Scheufele, and Shanto Iyengar (2016). "The end of framing as we know it ... and the future of media effects". *Mass Communication and Society 19*(1), 7–23.

Card, David, Alexandre Mas, Enrico Moretti, and Emmanuel Saez (2012). "Inequality at work: the effect of peer salaries on job satisfaction". *American Economic Review 102*(6), 2981–3003.

Chiswick, Barry R. and Paul W. Miller (1998). "English Language fluency among immigrants in the United States". *Research in Labor Economics 17*, 151–200.

– (2006). "Chapter 5 - Language skills and immigrant adjustment: the role of immigration policy". In: *Public Policy and Immigrant Settlement*. Ed. by Deborah A. Cobb-Clark and Siew-Ean Khoo. Cheltenham: Edward Elgar Publishing.

– (2007). *The economics of language: international analyses*. Oxfordshire: Routledge.

– (2015). "Chapter 5 - International migration and the economics of language". In: *Handbook of the Economics of International Migration*. Ed. by Barry R. Chiswick and Paul W. Miller. Amsterdam: North-Holland Publishing.

Chopra, Felix, Ingar Haaland, and Christopher Roth (2019). "Do people value more informative news?" *SSRN Scholarly Paper 3342595*, Social Science Research Network, Rochester (New York).

Cruces, Guillermo, Ricardo Perez-Truglia, and Martin Tetaz (2013). "Biased perceptions of income distribution and preferences for redistribution: Evidence from a survey experiment". *Journal of Public Economics 98*, 100–112.

Cullen, Zoë and Ricardo Perez-Truglia (2018). "How much does your boss make? The effects of salary comparisons". *NBER Working Paper 24841*, National Bureau of Economic Research, Cambridge (Massachusetts).

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

Dustmann, Christian and Arthur Van Soest (2002). "Language and the earnings of immigrants". *Industrial and Labor Relations Review 55*(3), 473–492.

Elias, Julio J., Nicola Lacetera, and Mario Macis (2015). "Sacred values? The effect of information on attitudes toward payments for human organs". *American Economic Review 105*(5), 361–365.

Friedrich Schiller University (2018). *Studierendenstatistik, Studierende und Studienfälle WiSe 2018/19 (Stichtag: 30.11.2018)*. Jena.

Geis-Thöne, Wido (2019). "Sprachkenntnisse entscheidend für die Arbeitsmarktintegration". *IW-Trends - Vierteljahresschrift zur empirischen Wirtschaftsforschung 46*(3), 73–89.

German Federal Ministry for Economic Affairs and Energy (2017). *Study and Work - Regionale Netzwerke zur Bindung von internationalen Studierenden*. Berlin.

German Federal Statistical Office (2019). *Bildung und Kultur - Studierende an Hochschulen - Endgültige Ergebnisse 2018/19*. Wiesbaden.

– (2020). *Data on the type of degree of international students - Email correspondence*. Wiesbaden.

Giesecke, Matthias and Eric Schuss (2019). "Heterogeneity in marginal returns to language training of immigrants". Ruhr Economic Working Paper 812, Ruhr Graduate School in Economics, Bochum.

Guven, Cahit and Asadul Islam (2015). "Age at migration, language proficiency, and socioeconomic outcomes: evidence from Australia". *Demography 52*(2), 513–542.

Haaland, Ingar, Christopher Roth, and Johannes Wohlfart (2021). "Designing information provision experiments". *Journal of Economic Literature*, forthcoming.

Isphording, Ingo E. (2015). "What drives the language proficiency of immigrants?" *IZA World of Labor 177*.

Isphording, Ingo E. and Sebastian Otten (2013). "The costs of Babylon—Linguistic distance in applied economics". *Review of International Economics 21*(2), 354–369.

Jacob, Brian A. and Tamara Wilder Linkow (2011). "Chapter 7 - Educational expectations and attainment". In: *Whither Opportunity?: Rising Inequality, Schools, and Children's Life Chances*. Ed. by Greg J. Duncan and Richard J. Murnane. New York: Russell Sage Foundation.

Koenings, Fabian, Tina Haussen, Stefan Toepfer, and Silke Uebelmesser (2021). "Coming to stay or to go? Stay intention and involved uncertainty of international students". *Journal of Regional Science 61*(2), 329–351.

Kuziemko, Ilyana, Michael I. Norton, Emmanuel Saez, and Stefanie Stantcheva (2015). "How elastic are preferences for redistribution? Evidence from randomized survey experiments". *American Economic Review 105*(4), 1478–1508.

Lergetporer, Philipp, Katharina Werner, and Ludger Woessmann (2020). "Educational inequality and public policy preferences: evidence from representative survey experiments". *Journal of Public Economics 188*, 104226.

Lochmann, Alexia, Hillel Rapoport, and Biagio Speciale (2019). "The effect of language training on immigrants' economic integration: empirical evidence from France". *European Economic Review 113*, 265–296.

Pineda, Jesús (2018). *Problemlagen und Herausforderungen internationaler Studierender in Deutschland - Ergebnisse einer qualitativen Vorstudie im Rahmen des Sesaba-Projekts.* Deutscher Akademischer Austausch Dienst (DAAD): Bonn.

Piopiunik, Marc, Franziska Kugler, and Ludger Wößmann (2017). "Einkommenserträge von Bildungsabschlüssen im Lebensverlauf: Aktuelle Berechnungen für Deutschland". *ifo Schnelldienst 70*(07), 19–30.

Raijman, Rebeca (2013). "Linguistic assimilation of first-generation Jewish South African immigrants in Israel". *Journal of International Migration and Integration 14*(4), 615–636.

SVR (2021). *The 8 core messages of the Annual Report 2021 | Diversity as the new normal? How Germany deals with difference as an immigration country.* Sachverständigenrat für Integration und Migration (SVR): Berlin.

Uebelmesser, Silke and Severin Weingarten (2017). "A macro-level analysis of adult-age language learning". *SSRN Scholarly Paper 2999688*, Social Science Research Network, Rochester (New York).

Wisniewski, Katrin (2018). "Sprache und Studienerfolg von Bildungsausländerinnen und -ausländern: Eine Längsschnittstudie an den Universitäten Leipzig und Würzburg". *Informationen Deutsch als Fremdsprache 45*(4), 573–597.

A. Appendix A: Main Appendix

Table A1: Two-sided t-Test on Difference in Means

	Mean (All)	Observations (All)	Mean (Control)	Mean (Treatment)	Difference (T. vs. C.)
Control Variables					
Age	23.231	324	23.650	22.807	0.843**
Female	0.627	324	0.595	0.658	−0.063
Region					
Africa	0.034	324	0.043	0.025	0.018
Americas	0.071	324	0.074	0.068	0.005
Asia	0.525	324	0.528	0.522	0.006
Europe	0.370	324	0.356	0.385	−0.029
Having a partner	0.145	324	0.153	0.137	0.017
Risk aversion	5.448	324	5.160	5.739	−0.580*
Patience	7.528	324	7.521	7.534	−0.013
Self-assessed German proficiency	2.750	324	2.748	2.752	−0.003
Linguistic barriers	3.173	324	3.129	3.217	−0.089
Social contact is important	4.429	324	4.417	4.441	−0.024
German course recommended	0.216	324	0.215	0.217	−0.003
Erasmus or similar	0.410	324	0.362	0.460	−0.098*
Bachelor program	0.302	324	0.270	0.335	−0.065
Master program	0.485	324	0.546	0.422	0.124**
Staying with certainty	2.540	324	2.638	2.441	0.197
Further Pre-Treatment Variables					
Learned German in home country	0.679	324	0.656	0.702	−0.045
Learned German in Germany	0.340	324	0.380	0.298	0.082
Framework German learned:					
In family	0.034	324	0.031	0.037	−0.007
At school	0.398	324	0.356	0.441	−0.085
Language course	0.509	324	0.534	0.484	0.049
Other	0.151	324	0.141	0.161	−0.020
Reason for no German course:					
No suitable courses	0.059	324	0.067	0.050	0.018
Not enough time	0.108	324	0.098	0.118	−0.020
Courses too expensive	0.065	324	0.067	0.062	0.005
Considered not necessary	0.071	324	0.061	0.081	−0.019
Does not apply	0.093	324	0.104	0.081	0.024
Other	0.006	324	0.000	0.012	−0.012

Two-sided t-Test on Difference in Means (Cont.)

	Mean (All)	Observations (All)	Mean (Control)	Mean (Treatment)	Difference (T. vs. C.)
Been in Germany previous to studies	0.599	324	0.620	0.578	0.042
Self-assessment relative to fellow Bachelor students	2.366	205	2.330	2.411	−0.081
German-speaking study program	0.559	324	0.577	0.540	0.036
Planned further education after graduation	2.087	323	2.012	2.161	−0.149
Degree of study content useful in home country	2.246	313	2.277	2.215	0.062
Planned graduation at the current university	0.594	313	0.574	0.614	−0.040
Self-assessed English proficiency	3.918	318	3.914	3.923	−0.008
Field of study					
Humanities	0.253	324	0.288	0.217	0.071
Law, Economics, Social Sciences	0.231	324	0.209	0.255	−0.046
Mathematics, Natural Sciences	0.235	324	0.233	0.236	−0.003
Medicine/Health Sciences	0.056	324	0.049	0.062	−0.013
Other (Art, Agriculture, Engineering, Sport)	0.043	324	0.049	0.037	0.012
No answer	0.182	324	0.172	0.193	−0.021
Opinion on learning German:					
Investment in future	4.495	321	4.543	4.447	0.097
Useful for studies	4.378	320	4.456	4.300	0.156
Helps with social integration	4.461	321	4.612	4.311	0.302***
Easier life in Germany	4.267	318	4.381	4.152	0.229**
More opport. in future	4.641	320	4.694	4.588	0.106
Improvement prof. opport.	4.642	321	4.696	4.588	0.108

The control group consists of 163 individuals and the treatment group of 161. Age ranges from 17 to 41. The self-assessed risk aversion and patience are measured using an 11-point scale. The self-assessed German proficiency ranges from No knowledge (0) to Very good knowledge (5). The variable 'Staying with certainty' is the product of the binary variable indicating to stay in Germany after graduation (1) or not (0) and a 7-point scale for the certainty on having the intention to stay. The expected linguistic barriers as well as importance of social contact are measured on a 5-point scale. 'Self-assessment relative to fellow Bachelor students' is measured on a 7-point scale ranging from Very good (1) to Very bad (7). 'Planned further education after graduation' is measured on a 3-point scale with No (0), Yes with a Master degree (1) and Yes, with a PhD (2). 'Degree of study content useful in home country' is measured on a 4-point scale ranging from Not at all (0) to Complete (3). The self-assessed English proficiency ranges from No knowledge (0) to Native Speaker (6). The answers on the 'Opinion on Learning German'-questions range from Strongly disagree (1) to Strongly agree (5). All other variables are measured on a binary/ categorical scale. For 324 observations information on the intention to participate in a German language course and the control variables are available. Due to item-non-responses the number of observations for some further pre-treatment variables are smaller. The stars indicate the following: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A2: Descriptive statistics (shares): International students in first semester (for winter semester 2018)

	(1) Sample	(2) University	(3) Germany
Female	0.627	0.574	0.502
Age (mean)	23.2	24.1	22.3
Origin region			
Europe	0.370	0.265	0.463
America	0.071	0.060	0.088
Africa	0.034	0.097	0.065
Asia	0.525	0.574	0.379
Australia and Oceania	0.000	0.002	0.003
Type of degree			
Bachelor	0.302	0.233	0.323
Master	0.485	0.660	0.315
Others/No information	0.213	0.107	0.315
Field of study [*]			
Humanities	0.314	0.187	0.183
Law, Economics, Social Sciences	0.287	0.332	0.276
Mathematics, Natural Sciences	0.292	0.371	0.164
Medicine, Health Sciences	0.069	0.062	0.043
Engineering Sciences	0.038	0.049	0.334
Observations	324	514	73801

The table displays the shares of all observations. [*] The 'Others/ No information' category is neglected.[18] In column (1) the numbers of observations corresponds to the sample used in the analyses. In column (2) the number of observations contain all international students without guest students. The number of observations in column (3) corresponds to the number of all international students enrolled at universities in Germany. Sources: Friedrich Schiller University (2018), German Federal Statistical Office (2019), and German Federal Statistical Office (2020). Own computations.

---

[18] In this descriptive comparison, I neglect the 'Others/ No information' category because a systematic bias between the fields of study and the missing information is unlikely and the shares for other study fields are relatively very small. The estimation analyses encompass this category.

Table A3: Two-sided t-Test on Difference in Means for English-questionnaire sample

	Mean (All)	Observations (All)	Mean (Control)	Mean (Treatment)	Difference (T. vs. C.)
Control Variables					
Age	22.904	167	22.963	22.849	0.114
Female	0.545	167	0.531	0.558	−0.027
Region					
Africa	0.042	167	0.049	0.035	0.014
Americas	0.090	167	0.099	0.081	0.017
Asia	0.521	167	0.506	0.535	−0.029
Europe	0.347	167	0.346	0.349	−0.003
Having a partner	0.138	167	0.148	0.128	0.020
Risk aversion	4.808	167	4.556	5.047	−0.491
Patience	7.976	167	8.247	7.721	0.526
Self-assessed German proficiency	1.731	167	1.679	1.779	−0.100
Social contact is important	4.377	167	4.395	4.360	0.035
Linguistic barriers	2.802	167	2.778	2.826	−0.048
German course recommended	0.401	167	0.432	0.372	0.060
Erasmus or similar	0.515	167	0.481	0.547	−0.065
Bachelor program	0.281	167	0.284	0.279	0.005
Master program	0.533	167	0.568	0.500	0.068
Staying with certainty	1.886	167	1.840	1.930	−0.091
Further Pre-Treatment Variables					
Learned German in my home country	0.623	167	0.605	0.640	−0.035
Learned German in Germany	0.198	167	0.198	0.198	−0.000
Framework German learned:					
In family	0.024	167	0.000	0.047	−0.047**
At school	0.383	167	0.358	0.407	−0.049
Language course	0.401	167	0.407	0.395	0.012
Other	0.084	167	0.049	0.116	−0.067
Reason for no German course:					
No suitable courses	0.072	167	0.099	0.047	0.052
Not enough time	0.156	167	0.148	0.163	−0.015
Courses too expensive	0.090	167	0.086	0.093	−0.007
Considered not necessary	0.066	167	0.062	0.070	−0.008
Does not apply	0.066	167	0.049	0.081	−0.032
Other	0.012	167	0.000	0.023	−0.023

Two-sided t-Test on Difference in Means for English-questionnaire sample (Cont.)

	Mean (All)	Observations (All)	Mean (Control)	Mean (Treatment)	Difference (T. vs. C.)
Been in Germany previous to studies	0.437	167	0.481	0.395	0.086
Self-assessment relative to fellow Bachelor students	2.310	100	2.212	2.417	−0.205
German-speaking study program	0.305	167	0.346	0.267	0.078
Planned further education after graduation	2.139	166	2.050	2.221	−0.171
Degree of study content useful in home country	2.292	161	2.351	2.238	0.113
Planned degree at the current university	0.610	164	0.613	0.607	0.005
Self-assessed English proficiency	4.176	165	4.148	4.202	−0.054
Field of study					
Humanities	0.174	167	0.247	0.105	0.142**
Law, Economics, Social Sciences	0.216	167	0.185	0.244	−0.059
Mathematics, Natural Sciences	0.359	167	0.358	0.360	−0.002
Medicine/Health Sciences	0.018	167	0.012	0.023	−0.011
Other (Art, Agriculture, Engineering, Sport)	0.006	167	0.000	0.012	−0.012
No answer	0.228	167	0.198	0.256	−0.058
Opinion on learning German					
Investment in future	4.343	166	4.432	4.259	0.173
Useful for studies	4.012	165	4.088	3.941	0.146
Helps with social integration	4.440	166	4.550	4.337	0.213
Easier life in Germany	4.277	166	4.395	4.165	0.230
More opport. in future	4.575	167	4.716	4.442	0.274**
Improvement prof. opport.	4.551	167	4.642	4.465	0.177

The control group consists of 81 individuals and the treatment group of 86. Age ranges from 19 to 40. The self-assessed risk aversion and patience are measured using an 11-point scale. The self-assessed German proficiency ranges from No knowledge (0) to Very good knowledge (5). The variable 'Staying with certainty' is the product of the binary variable indicating to stay in Germany after graduation (1) or not (0) and a 7-point scale for the certainty on having the intention to stay. The expected linguistic barriers as well as importance of social contact are measured on a 5-point scale. 'Self-assessment relative to fellow Bachelor students' is measured on a 7-point scale ranging from Very good (1) to Very bad (7). 'Planned further education after graduation' is measured on a 3-point scale with No (0), Yes with a Master degree (1) and Yes, with a PhD (2). 'Degree of study content useful in home country' is measured on a 4-point scale ranging from Not at all (0) to Complete (3). The self-assessed English proficiency ranges from No knowledge (0) to Native Speaker (6). The answers on the 'Opinion on Learning German'-questions range from Strongly disagree (1) to Strongly agree (5). All other variables are measured on a binary/ categorical scale. For 167 observations information on the intention to participate in a German language course and the control variables are available. Due to item-non-responses the number of observations for some further pre-treatment variables are smaller. The stars indicate the following: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A4: Two-sided t-Test on Difference in Means with New Control Group for English-questionnaire sample

	Mean (All)	Observations (All)	Mean (Control)	Mean (Treatment)	Difference (T. vs. C.)
Control Variables					
Age	23.030	201	23.207	22.762	0.444
Female	0.587	201	0.579	0.600	−0.021
Region					
Africa	0.055	201	0.066	0.037	0.029
Americas	0.000	201	0.000	0.000	0.000
Asia	0.562	201	0.545	0.588	−0.042
Europe	0.383	201	0.388	0.375	0.013
Having a partner	0.124	201	0.116	0.138	−0.022
Risk aversion	5.164	201	5.223	5.075	0.148
Patience	7.726	201	7.760	7.675	0.085
Self-assessed German proficiency	1.741	201	1.769	1.700	0.069
Social contact is important	4.343	201	4.339	4.350	−0.011
Linguistic barriers	2.896	201	2.942	2.825	0.117
German course recommended	0.333	201	0.314	0.362	−0.048
Erasmus or similar	0.443	201	0.364	0.563	−0.199***
Bachelor program	0.209	201	0.165	0.275	−0.110*
Master program	0.572	201	0.628	0.487	0.141**
Staying with certainty	1.945	201	2.091	1.725	0.366
Further Pre-Treatment Variables					
Learned German in my home country	0.577	201	0.529	0.650	−0.121*
Learned German in Germany	0.214	201	0.231	0.188	0.044
Framework German learned:					
In family	0.025	201	0.008	0.050	−0.042*
At school	0.338	201	0.298	0.400	−0.102
Language course	0.408	201	0.421	0.388	0.034
Other	0.095	201	0.074	0.125	−0.051
Reason for no German course:					
No suitable courses	0.075	201	0.091	0.050	0.041
Not enough time	0.214	201	0.248	0.163	0.085
Courses too expensive	0.090	201	0.083	0.100	−0.017
Considered not necessary	0.055	201	0.041	0.075	−0.034
Does not apply	0.075	201	0.074	0.075	−0.001
Other	0.055	201	0.074	0.025	0.049

Two-sided t-Test on Difference in Means with New Control Group for English- questionnaire sample (Cont.)

	Mean (All)	Observations (All)	Mean (Control)	Mean (Treatment)	Difference (T. vs. C.)
Been in Germany previous to studies	0.423	201	0.438	0.400	0.038
Self-assessment relative to fellow Bachelor students	2.236	127	2.061	2.556	−0.495***
German-speaking study program	0.264	201	0.264	0.263	0.002
Planned further education after graduation	1.898	147	1.493	2.237	−0.745***
Degree of study content useful in home country	2.387	181	2.495	2.244	0.252*
Planned degree at the current university	0.656	195	0.695	0.597	0.098
Self-assessed English proficiency	4.136	198	4.175	4.077	0.098
Field of study					
Humanities	0.137	190	0.155	0.113	0.042
Law, Economics, Social Sciences	0.258	190	0.273	0.237	0.035
Mathematics, Natural Sciences	0.368	190	0.382	0.350	0.032
Medicine/Health Sciences	0.016	190	0.009	0.025	−0.016
Other (Art, Agriculture, Engineering, Sport)	0.021	190	0.027	0.013	0.015
No answer	0.200	190	0.155	0.263	−0.108*
Opinion on learning German					
Investment in future	4.492	199	4.667	4.228	0.439***
Useful for studies	4.055	199	4.167	3.886	0.281
Helps with social integration	4.507	201	4.653	4.287	0.365***
Easier life in Germany	4.332	199	4.467	4.127	0.340**
More opport. in future	4.653	199	4.824	4.400	0.424***
Improvement prof. opport.	4.575	200	4.675	4.425	0.250**

The control group consists of 121 individuals and the treatment group of 80. Age ranges from 18 to 40. The self-assessed risk aversion and patience are measured using an 11-point scale. The self-assessed German proficiency ranges from No knowledge (0) to Very good knowledge (5). The variable 'Staying with certainty' is the product of the binary variable indicating to stay in Germany after graduation (1) or not (0) and a 7-point scale for the certainty on having the intention to stay. The expected linguistic barriers as well as importance of social contact are measured on a 5-point scale. 'Self-assessment relative to fellow Bachelor students' is measured on a 7-point scale ranging from Very good (1) to Very bad (7). 'Planned further education after graduation' is measured on a 3-point scale with No (0), Yes with a Master degree (1) and Yes, with a PhD (2). 'Degree of study content useful in home country' is measured on a 4-point scale ranging from Not at all (0) to Complete (3). The self-assessed English proficiency ranges from No knowledge (0) to Native Speaker (6). The answers on the 'Opinion on Learning German'-questions range from Strongly disagree (1) to Strongly agree (5). All other variables are measured on a binary/ categorical scale. For 201 observations information on the intention to participate in a German language course and the control variables are available. Due to item-non-responses the number of observations for some further pre-treatment variables are smaller. The stars indicate the following: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.
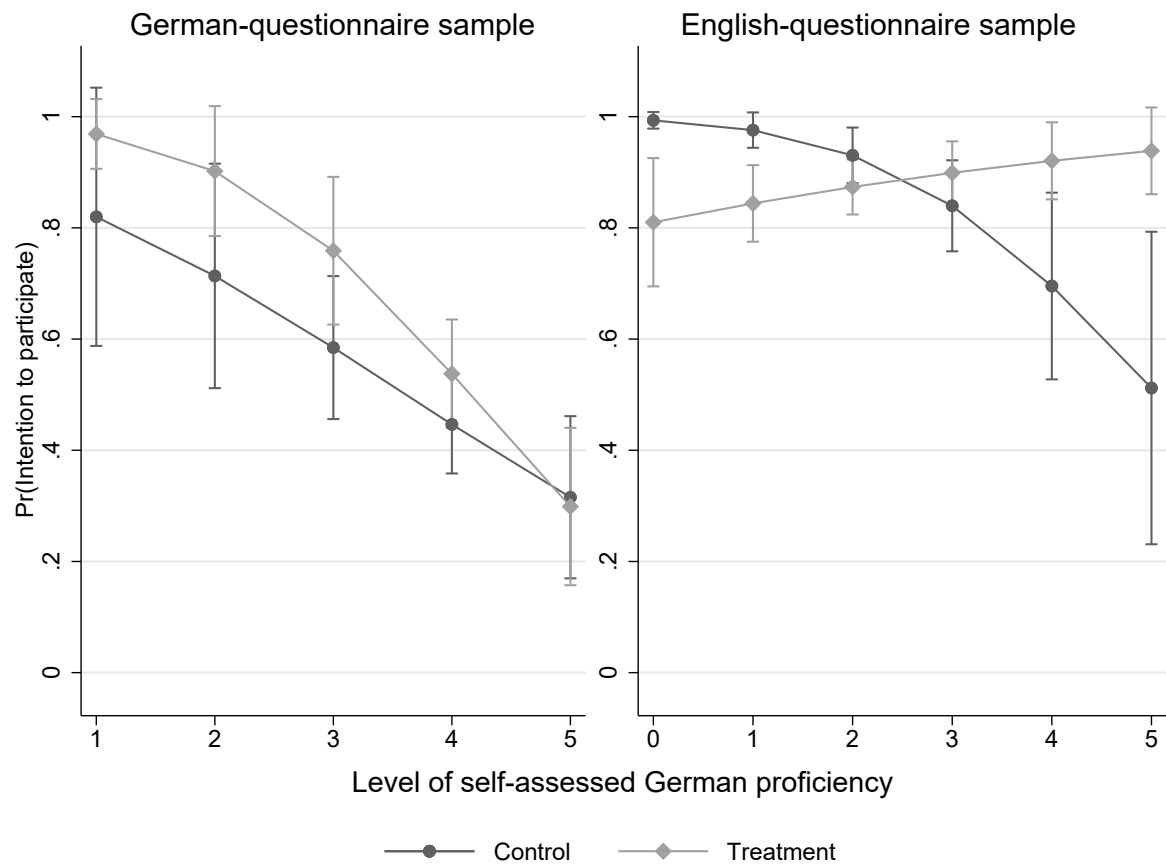
Figure A1: Level analysis by self-assessed German proficiency

Notes: The left-hand side displays the predicted propensities to have the intention to participate with and without treatment for each level of self-assessed German proficiency in the German-questionnaire sample. The right-hand side displays the predicted propensities to have the intention to participate with and without treatment for each level of self-assessed German proficiency in the English-questionnaire sample. For the German-questionnaire sample there is no value for a self-assessed German proficiency level of 0 since there are no observations in this category. Calculations are based on the estimations shown in columns (2) and (4) of Table 3. The plots are displayed with 90% CIs.

Table A5: Heterogeneous treatment effects by levels of self-assessed English proficiency

	(1) All questionnaires	(2) All questionnaires: With control variables	(3) English questionnaires	(4) English questionnaires: With control variables
Treatment	0.092	0.007	−2.052	−1.122
	(0.534)	(0.626)	(1.313)	(1.510)
English prof.	−0.011	−0.055	−0.400	−0.637**
	(0.099)	(0.123)	(0.258)	(0.298)
Treatment × English prof.	−0.003	0.019	0.464	0.341
	(0.135)	(0.159)	(0.301)	(0.346)
Control variables		Yes		Yes
Observations	300	300	151	151
Pseudo-R^2	0.001	0.236	0.030	0.291

The dependent variable is the intended participation in a German course. The control variables include age, gender, region of origin, an indicator for having a partner, the level of risk aversion, of patience, and the level of self-assessed German proficiency. They further include indications on whether linguistic barriers are expected, whether social contact is important, whether a German course is recommended for the study program, being part of Erasmus or a similar program, being enrolled in a Bachelor or Master program, and the intention to stay in Germany after graduation multiplied by the level of certainty. Column (4) does not control for the region of origin as some observations would be dropped from the sample and, hence, the numbers of observations would not align with column (3). However, controlling for the region of origin also in this column does not affect the result. For the analyses displayed respondents with English as mother tongue are excluded. Standard errors are displayed in parentheses. The stars indicate the following: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

B. Appendix B: Logit and LPM estimations

Table B1: Treatment effect with and without controls – LOGIT

	(1)		(2) with control variables	
Treatment	0.130	(0.243)	0.172	(0.319)
Control variables			Yes	
Observations	324		324	
Pseudo-R^2	0.001		0.310	

The dependent variable is the intended participation in a German course. The control variables include age, gender, region of origin, an indicator for having a partner, the level of risk aversion, of patience, and of self-assessed German proficiency. They further include indications on whether linguistic barriers are expected, whether social contact is important, whether a German course is recommended for the study program, being part of Erasmus or a similar program, being enrolled in a Bachelor or Master program, and the intention to stay in Germany after graduation multiplied by the level of certainty. Standard errors are displayed in parentheses. The stars indicate the following: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B2: Treatment effect with and without controls – LPM

	(1)		(2) with control variables	
Treatment	0.027	(0.051)	0.026	(0.044)
Control variables			Yes	
Observations	324		324	
R^2	0.001		0.328	

The dependent variable is the intended participation in a German course. The control variables include age, gender, region of origin, an indicator for having a partner, the level of risk aversion, of patience, and of self-assessed German proficiency. They further include indications on whether linguistic barriers are expected, whether social contact is important, whether a German course is recommended for the study program, being part of Erasmus or a similar program, being enrolled in a Bachelor or Master program, and the intention to stay in Germany after graduation multiplied by the level of certainty. Standard errors are displayed in parentheses. The stars indicate the following: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B3: Heterogeneous treatment effects by self-assessed German proficiency and German questionnaire – LOGIT

	(1) Proficiency	(2) Proficiency with control variables	(3) German questionnaire	(4) German questionnaire with control variables
Treatment	−1.372	−1.158	−0.291	0.101
	(0.889)	(0.914)	(0.493)	(0.540)
German prof.	−1.126***	−0.866***		
	(0.194)	(0.226)		
Treatment × German prof.	0.448*	0.394		
	(0.240)	(0.252)		
German questionnaire			−2.309***	−1.113**
			(0.433)	(0.525)
Treatment × German questionnaire			0.523	0.070
			(0.588)	(0.657)
Control variables		Yes		Yes
Observations	324	324	324	324
Pseudo-R^2	0.224	0.317	0.153	0.280

The dependent variable is the intended participation in a German course. The control variables include age, gender, region of origin, an indicator for having a partner, the level of risk aversion, and of patience. They further include indications on whether linguistic barriers are expected, whether social contact is important, whether a German course is recommended for the study program, being part of Erasmus or a similar program, being enrolled in a Bachelor or Master program, and the intention to stay in Germany after graduation multiplied by the level of certainty. Standard errors are displayed in parentheses. The stars indicate the following: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B4: Heterogeneous treatment effects by self-assessed German proficiency levels and German questionnaire – LPM

	(1) Proficiency	(2) Proficiency with control variables	(3) German questionnaire	(4) German questionnaire with control variables
Treatment	−0.105	−0.070	−0.029	0.003
	(0.090)	(0.087)	(0.065)	(0.062)
German prof.	−0.164***	−0.114***		
	(0.020)	(0.024)		
Treatment × German prof.	0.048*	0.036		
	(0.028)	(0.028)		
German questionnaire			−0.426***	−0.206***
			(0.065)	(0.076)
Treatment × German questionnaire			0.087	0.042
			(0.093)	(0.091)
Control variables		Yes		Yes
Observations	324	324	324	324
R^2	0.238	0.331	0.177	0.298

The dependent variable is the intended participation in a German course. The control variables include age, gender, region of origin, an indicator for having a partner, the level of risk aversion, and of patience. They further include indications on whether linguistic barriers are expected, whether social contact is important, whether a German course is recommended for the study program, being part of Erasmus or a similar program, being enrolled in a Bachelor or Master program, and the intention to stay in Germany after graduation multiplied by the level of certainty. Standard errors are displayed in parentheses. The stars indicate the following: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B5: Heterogeneous treatment effects by self-assessed German proficiency: Different samples and robustness checks– LOGIT

	(1) German questionnaires: Proficiency	(2) German questionnaires: Proficiency with control. variables	(3) English questionnaires: Proficiency	(4) English questionnaires: Proficiency with control variables	(5) English questionnaires: New control group group	(6) English questionnaires: New control group with control variables	(7) English questionnaires: German program	(8) English questionnaires: German program with control variables
Treatment	1.150	2.630	−2.650**	−3.118**	−2.154**	−3.469***	−1.311	−1.031
	(1.684)	(1.993)	(1.228)	(1.528)	(0.977)	(1.184)	(0.814)	(0.931)
German prof.	−0.868***	−0.697*	−1.084***	−1.157***	−0.870***	−1.210***		
	(0.290)	(0.364)	(0.334)	(0.422)	(0.241)	(0.318)		
Treatment×German prof.	−0.202	−0.575	0.952**	1.430***	0.685**	1.242***		
	(0.419)	(0.488)	(0.405)	(0.540)	(0.333)	(0.424)		
German progr.							−1.939**	−1.824
							(0.855)	(1.139)
Treatment×German progr.							1.909*	2.996**
							(1.122)	(1.368)
Control variables		Yes		Yes		Yes		Yes
Observations	157	157	167	167	201	201	167	167
Pseudo-R^2	0.129	0.262	0.124	0.343	0.120	0.273	0.054	0.293

The first to columns use the German-questionnaire sample and the succeeding columns use the English-questionnaire sample. The dependent variable is the intended participation in a German course. The control variables include age, gender, region of origin, an indicator for having a partner, the level of risk aversion, and of patience. They further include indications on whether linguistic barriers are expected, whether social contact is important, whether a German course is recommended for the study program, being part of Erasmus or a similar program, being enrolled in a Bachelor or Master program, and the intention to stay in Germany after graduation multiplied by the level of certainty. Columns (4) and (8) do not control for the region of origin as the seven African students, who all want participate in a German course, would be dropped from the sample and, hence, the numbers of observations would not align with columns (3) and (7), respectively. However, controlling for the region of origin also in these columns does not affect the results. Standard errors are displayed in parentheses. The stars indicate the following: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B6: Heterogeneous treatment effects by self-assessed German proficiency: Different samples and robustness checks– LPM

	(1) German questionnaires: Proficiency	(2) German questionnaires: Proficiency with control. variables	(3) English questionnaires: Proficiency	(4) English questionnaires: Proficiency with control variables	(5) English questionnaires: New control group group	(6) English questionnaires: New control group with control variables	(7) English questionnaires: German program	(8) English questionnaires: German program with control variables
Treatment	0.146	0.396	−0.172**	−0.155**	−0.146**	−0.184**	−0.089	−0.060
	(0.296)	(0.301)	(0.078)	(0.077)	(0.072)	(0.076)	(0.059)	(0.058)
German prof.	−0.189***	−0.122**	−0.102***	−0.090***	−0.083***	−0.087***		
	(0.055)	(0.061)	(0.027)	(0.027)	(0.020)	(0.023)		
Treatment×German prof.	−0.020	−0.086	0.086**	0.094***	0.060*	0.074**		
	(0.075)	(0.075)	(0.036)	(0.036)	(0.033)	(0.034)		
German progr.							−0.177**	−0.127
							(0.074)	(0.087)
Treatment×German progr.							0.173	0.205*
							(0.107)	(0.104)
Control variables		Yes		Yes		Yes		Yes
Observations	157	157	167	167	201	201	167	167
R^2	0.162	0.308	0.086	0.223	0.084	0.163	0.036	0.186

The first to columns use the German-questionnaire sample and the succeeding columns use the English-questionnaire sample. The dependent variable is the intended participation in a German course. The control variables include age, gender, region of origin, an indicator for having a partner, the level of risk aversion, and of patience. They further include indications on whether linguistic barriers are expected, whether social contact is important, whether a German course is recommended for the study program, being part of Erasmus or a similar program, being enrolled in a Bachelor or Master program, and the intention to stay in Germany after graduation multiplied by the level of certainty. Columns (4) and (8) do not control for the region of origin as the seven African students, who all want participate in a German course, would be dropped from the sample and, hence, the numbers of observations would not align with columns (3) and (7), respectively. However, controlling for the region of origin also in these columns does not affect the results. Standard errors are displayed in parentheses. The stars indicate the following: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.