Communication Effort and the Cost of Language: Evidence from Stack Overflow

Jacopo Bregolin*
University of Liverpool

January 2022

Abstract

The transmission of information is crucial for productivity and growth, but language differences may limit its effectiveness. In this paper, I empirically investigate how much communication effort is affected by the exogenous cost of language, and how this effect interacts with the incentives between sender and receiver. I exploit the introduction of websites for languages different from English on a question-and-answering platform and compare the behavior of non-English speaking users before and after the introduction. Results are consistent with a simple framework of pairwise communication and show that the quality of communication improves when writers use their first language, rather than English. In addition, the size of the effect increases on users' cost of using English and on the effort of the questioner. Finally, the increase in senders' effort after a drop in the cost of language is larger when the sender is more incentivised. The paper sheds lights on the trade-off that the platform faces when choosing whether to adopt additional languages.

JEL Codes: D82, D83, Z13

Keywords: Cost of language, information transmission, knowledge platforms

*email: jacopo.bregolin@liverpool.ac.uk. This paper is the second chapter of my PhD dissertation. I am extremely indebted to my advisor, Milo Bianchi, co-advisor, Daniel L. Chen, and committee members Jacques Crémer and Daniel Ershov. Their support and guidance have been invaluable. I also thank April Franco, Astrid Hopfensitz, and Marta Troya-Martinez for helpful discussions and feedback. I thank seminar participants at the University of Liverpool Management School internal seminar for useful comments.

1 Introduction

Complex languages and the transmission of information have been identified as crucial factors for human evolution, potentially being the main source of our differentiation from the animal world (Diamond 1991). The sharing of knowledge allows people to take advantage of each others' human capital investments, speeding up learning and productivity. Nevertheless, information transmission may be limited in several ways. On one side, communication may be affected by the incentives of the information holder. On the other, exogenous cognitive boundaries may constrain our ability to share information.[1]

In this paper I focus on the latter constraint, and, in particular, I study to what extent the use of a foreign language affects effort choices in communication. This is a major concern in all contexts where individuals do not share the same language, but organizations or institutions still want to maximize information sharing (Ginsburgh and Weber 2011). In the digital era, a leading example is provided by knowledge platforms that aim to aggregate information, like Wikipedia or Stack Overflow. These platforms aim to be global and face the challenging choice of using one or multiple languages.

To study this trade-off, I use data from Stack Overflow, a question-and-answer website, and exploit the staggered introduction of versions of the website that use languages different from English. This natural experiment allows me to measure effort choices of non-English speaking users before and after the introduction of the new site, that is before and after they were able to use their native language on top of English.

In addition, by observing variation in the degree of (virtual) monetary incentives between the users asking and answering the questions, I study how the impact of the exogenous cost of language interacts with the incentive system, bridging, in this way, the literature of strategic information transmission with the literature on bounded communication in teams.

The paper shows that users increase their communication effort when speaking in their native language. The effect is specifically relevant for users with a high cost of using English, it increases in the effort of the questioner, and in the degree to which the user is incentivized by the questioner.

The study of communication effort choices is particularly relevant in the context of Stack Overflow. Question-and-answer websites' success is strictly based on the quantity and quality of the information provided: the platform has then all incentives to reduce

[1] The economic literature has theoretically investigated both of these constraints, but generally focusing on one or the other. On one side, the literature has looked at incentives and strategic information transmission, either without costs of effort in communication, i.e. the *Cheap Talk* literature (Crawford and Sobel 1982, Austen-Smith and Banks 2000, Asher and Lascarides 2013, Sobel 2013, Dilmé 2016), or with strategic choice of effort (Dewatripont and Tirole 2005) as in the signalling literature (Spence (1973), Gambetta (2011)). On the other sides, the team-theory literature (Marschak and Radner (1972)) has focused on environments where incentives are perfectly aligned, but exogenous constraints affect the ability to communicate, for instance bounded cognitive abilities or costs in information processing (Arrow 1974, Bolton and Dewatripont 1994, Crémer, Garicano, and Prat 2007, Blume and Board 2013, Blume 2018, Dilmé 2018). In this paper, I put together the two strands and look at the interaction between incentives and exogenous costs.

the barriers to participation and effort provision in communication.[2]

The platform's optimal strategies are not trivial. By allowing users to communicate in languages different from English, the platform reduces its ability to aggregate information, and it may increase inexperienced users' participation.[3] On the other side, it reduces communication costs and barriers to participation, and it improves diversity.[4]

Empirically I do not find specific negative externalities on the original English platform from the introduction of non-English websites, which suggests that in practice these trade-offs may compensate each other.

Why do these trade-offs arise? How do users decide their effort provision? To illustrate in what ways exogenous communication costs and incentives affect effort choices, I propose a simple theoretical framework of pairwise communication that incorporates both factors. The framework is instrumental to show how variations in the exogenous cost of language may interact with the incentives, and how it may induce trade-offs at the platform level.

To test empirically the implication of the model, I exploit a natural experiment. Stack Overflow was created in 2008 in English. Nevertheless, the platform managers introduced additional websites in Russian, Portuguese, Japanese, and Spanish with the same purpose and functioning as the initial website. After recovering all contributions of users who participated in those non-English websites, I compare the effort provided before and after the introduction of their native-language website, where the effort is measured as the number of separate pieces of code in their answers.[5] As a control group, I include in the analysis contributions made by a random sample of users who have not joined the non-English websites.

Methodologically, I adopt the staggered difference-in-difference approach. I use the estimation technique developed by Borusyak, Jaravel, and Spiess (2021) and compare the results with the more standard Two-Way Fixed-Effects approach, which has anyway been proven biased by recent econometric literature (Callaway and Sant'Anna 2020, de Chaisemartin and D'Haultfœuille 2020, Sun and Abraham 2020, Borusyak et al. 2021). To understand how the effect of the reduction in exogenous communication costs may interact with other factors, I run separate difference-in-difference regression for different levels of the initial cost of using English, the effort put by the author of the

[2] StackOverflow is probably the main source of help for computer programmers. Solutions provided in the platform affect the code of programmers all over the world, and mistakes or bad information may have a large impact. It happened for instance that the most copied code snippet from StackOverflow had an error: https://programming.guide/worlds-most-copied-so-snippet.html.

[3] Bao, Hecht, Carton, Quaderi, Horn, and Gergle (2012) show that in Wikipedia some information is often available in some languages, but not in others. Stack Overflow may easily suffer from the same problem.

[4] These types of trade-offs are generalizable to any economic environment and all national institutions with language diversity face the critical choice of imposing or not a unique language (Ginsburgh and Weber (2011)). Opposite examples are France, where a unique language was imposed, and Spain, where different languages are still widely used.

[5] I include a robustness check where the measure of effort (i.e. quality in this context) is the likelihood that the answer is selected as the best answer, i.e. the answer that indeed provided the solution to the author of the question.

question, and the degree of virtual remuneration offered by the author of the question.

The analysis shows that users increase effort when they write in their native language, compared to when they write in English. The effect is mainly driven by users that have a high cost of using English, identified as the users that switched the most to the native-language website. The effect is larger when the author of the question puts more effort and increases when the author of the questions auctions a higher virtual reward.

To investigate the trade-offs of the platform, I finally measure the effect of the introduction of non-English websites on the average quality in the original English platform. While the different methodologies provide contradicting results, the preferred specification suggests that there are no significant externalities on the English website.

To my knowledge, this is the first paper that quantifies the effect of the cost of language on communication effort. Some experimental literature has tested communication games, with or without communication frictions (Lafky and Wilson 2020 and Blume, DeJong, Kim, and Sprinkle 2001 respectively, for instance). Battiston, Blanes I Vidal, and Kirchmaier (2021) also studies exogenous communication costs and their effect on performance. Their focus is anyway on communication frictions arising from not being able to talk face-to-face, rather than the language itself.

This paper finally contributes to the trade literature, which has investigated the negative effect of language barriers on trade (Melitz 2008, Lohmann 2011).

For what follows, section 2 discusses communication in Q&A websites and the case of Stack Overflow, section 3 presents a simple theoretical framework, section 4 presents the data, section 5 and 6 report the empirical results, and section 7 concludes.

2 Communication in Q&A platforms

Question and Answers websites are online platforms that allow users to ask new questions or answer existing ones. Examples of such websites are *Stack Overflow*, *Yahoo! Answers*[6], or *Quora*.

The content of these websites is particularly useful for the analysis of communication strategies for several reasons. First of all, they provide detailed data on information transmission, including the information requested and the information provided, both generally not observed. This richness allows measuring effort in information transmission on both sides of the communication. In addition, as communities tend to be large, interpersonal relationships may be weaker. As a consequence, communication strategies are less likely to be affected by unobserved factors, like friendship or long-term norms, very common within firms. Finally, question and answers websites allow the researcher to observe a very large number of communication interactions, allowing more flexible statistical analysis.

[6] *Yahoo! Answer* has shut down in April 2021

2.1 Stack Overflow

For the empirical investigation, the paper relies on data from Stack Overflow, a question and answers website that focuses on topics related to computer programming. Questions may concern, for instance, how to use programming languages for data analysis or software development, or how to solve coding bugs. The website has the objective to be the main resource of information for all possible problems that programmers may encounter.[7] Key features of the platform is that it is crowd-based and free of charge. In other words, any internet user who register (for free) on the website can ask questions and/or provide answers to other questions. Contributors are not remunerated.[8]

Stack Overflow stands out from other sites because of the size of its welfare impact: many programmers are self-learned and Stack Overflow provides a large community willing to help. As of June 2021 indeed, Stack Overflow receives more than one hundred million monthly visits.[9]. In addition, the quality of Stack Overflow's answers has an enormous societal impact, as a very large amount of programmers rely on them.[10]

2.2 Language used in Stack Overflow

As of today, there are five different websites of Stack Overflow, each using a different language, namely English, Russian, Japanese, Portuguese, and Spanish. Note that, apart form the language, their working is exactly identical. Each website anyway became public at different times. Stack Overflow was first launched in English in September 2008. The platform was implemented in English as the founders are Americans and the use of English is the norm in the programming community. Nevertheless, they realized that a significant part of the programming community would not be able or may have problems accessing English content. After some discussion, they decided to allow the opening of Stack Overflow in other languages than English.[11]

The platform designers chose those 4 additional languages on the ground that large communities of programmers speak them and, at the same time, they may not speak English. The introduction of each website followed some *beta* periods before the roll out of the final version.[12]

Figure 2.2 shows the timeline of the introduction of the different websites.

The case of Stack Overflow in Russian

The introduction of Stack Overflow in Russian followed a slightly different process. In 2010, some Russian programmers decided to create a clone of English Stack Overflow in Russian. They created a website called HashCode which was replicating Stack Overflow features and purpose. Once the company behind Stack Overflow decided to open a version in Russian, they acquired HashCode, and on March 31st 2015 all posts from

[7] https://www.joelonsoftware.com/2008/09/15/stack-overflow-launches/

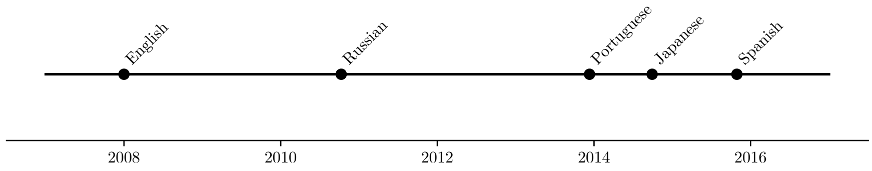
[8] Nevertheless, the platform has implemented several incentive systems, including virtual rewards.

[9] https://stackoverflow.com/company

[10] There are anecdotes that mistakes in Stack Overflow translated into long-term bugs in relevant and popular software

[11] `https://stackoverflow.blog/2014/02/13/cant-we-all-be-reasonable-and-speak-english/`

[12] Appendix ?? provides more details on the introduction of the websites



HashCode were imported in the Russian version of Stack Overflow. Formally then, the russian version of Stack Overflow appeared in 2015. Figure 2.2 reports the 2010 date as the data includes all the HashCode content.

3 Theoretical Illustration

Communications costs, intended as the difficulties in providing clear and informative solutions, may affect users' participation decisions, as well as the quality of their content. Nevertheless, it is not obvious in what ways these costs interact with other variables that may affect the effort decision, as the alignment of interest between communicating parties and the other party effort choice. In addition, it may not be trivial to understand the trade-off of the platform in providing the option to contribute in different languages.

In this section, I present a simple framework that models unilateral information transmission in the platform, that is, between who asks and who answers the questions[13]. In this simple environment, I abstract from strategic behavior and I assume that pairwise communication is independent of other communicating pairs[14]. In addition, I do not consider the incentive system implemented by the platform to affect the decision process, and I rely on user and time fixed-effects at the stage of the analysis to control for heterogeneous sensitivity to the incentive system.

Let Bob be a programmer that needs to understand how to implement some features in his software. After multiple attempts, he decides to ask his problem to the Stack Overflow community as, otherwise, he is not able to proceed on his project. Alice instead is a community member that sometimes answers questions on the platform.

Bob and Alice independently decide their communication strategies: it could be thought as they just keep the same strategies every time they participate in the website. This assumption is reasonable because the community is very large and both Bob and Alice cannot anticipate who will ask a question or provide the answer.

Alice then understands the solution to Bob's problem, decides whether to provide the answer and, in case, publishes her solution. Bob then implements the features thanks to Alice's help.

Note that the game is static as strategies are decided ex-ante.

[13]The modeling approach and the functional forms are inspired by Calvó-Armengol, de Martí, and Prat (2015) and the communication literature in Organizational Economics

[14]This assumption is justified by the fact that Stack Overflow is a very large community, and it is hard for users to have accurate beliefs over other users' decisions

More formally, let the information that Bob needs be θ, of which he only knows the ex-ante distribution:

$$\theta \sim \mathscr{N}\left(0, \frac{1}{s}\right).$$

In addition let, Bob's and Alice's efforts be defined, respectively, as E_Q and E_A, where E_Q captures the clarity and informativeness of the question, and E_A the clarity and informativeness of the answer. The cost of effort, C, depends on the cost of using a given language (λ) and the experience or general knowledge in the subject (k), and it is defined as:

$$C_Q = \frac{\lambda_Q}{k_Q}; \quad C_A = \frac{\lambda_A}{k_A}$$

for Bob and Alice's costs of effort respectively. The interpretation of the knowledge parameter is that the more the user is experienced, the more he can get to the point exactly, providing an accurate description of the question/solution. The crucial point that I want to capture is that the cost of language affects the intelligibility of the message, so that if a message is badly written it cannot be understood, while the experience affects the probability of misunderstanding, for example via misleading content.

Once Bob has published his questions, Alice provides the answer with the message m such that:

$$m = \theta + \varepsilon + \eta$$

where ε and η are noise terms that shrink with the agents' efforts. More precisely, $\varepsilon \sim \mathscr{N}\left(0, \frac{1}{E_Q}\right)$ and $\eta \sim \mathscr{N}\left(0, \frac{1}{E_A}\right)$.

Finally, let $a \in (-\infty, \infty)$ be what Bob will finally do to solve his problem.

The utility functions of Bob and Alice are, respectively,

$$\begin{aligned} U_Q &= -\left((a-\theta)^2 + C_Q^2 E_Q\right) \\ U_A &= -\left(\gamma(a-\theta)^2 + C_A^2 E_A\right). \end{aligned}$$

Bob wants to minimize any error in implementing the features in his software, and Alice will internalize Bob's utility to a certain degree $\gamma \in [0, 1]$.

3.1 Optimal effort strategies

For a given question, how much effort Alice will decide to make? Proceeding backward, Bob selects the action a^* such that:

$$\begin{aligned} & a^* \equiv \arg\max_a \mathbb{E}[-\left((a-\theta)^2 + C_Q^2 E_Q\right)|m] \\ \Longleftrightarrow\ & a^* \equiv \arg\max_a -a^2 - \mathbb{E}[\theta^2|m] + 2a\mathbb{E}[\theta|m] + C_Q^2 E_Q. \end{aligned}$$

The optimal action is then given by:

$$
\begin{aligned}
&-2a^* + 2\mathbb{E}[\theta|m] = 0 \\
&a^* = \mathbb{E}[\theta|m] = \beta m \quad \text{with} \quad \beta \equiv \frac{E_Q E_A}{E_Q E_A + E_Q s + E_A s}
\end{aligned}
$$

where the last equality holds because of Bayes Normal updating.

To find her optimal effort level, Alice solves:

$$
\max_{E_A \geq 0} \mathbb{E}[-\left(\gamma(a-\theta)^2 + C_A^2 E_A\right)]
$$

which, given the action expected to be chosen by Bob, rewrites as:

$$
\begin{aligned}
&\max_{E_A \geq 0} -\gamma \mathbb{E}[(\beta m - \theta)^2] - C_A^2 E_A \\
\iff &\max_{E_A \geq 0} -\gamma \mathbb{E}[\beta m - \theta]^2 + -\gamma \mathbb{V}[\beta m - \theta] - C_A^2 E_A \\
\iff &\max_{E_A \geq 0} -\gamma \mathbb{V}[\beta m - \theta] - C_A^2 E_A \\
\iff &\max_{E_A \geq 0} -\gamma \left(\beta^2 \frac{1}{s} + \beta^2 \frac{1}{E_A} + \beta^2 \frac{1}{E_Q} + \frac{1}{s} - 2\beta \frac{1}{s}\right) - C_A^2 E_A \\
\iff &\max_{E_A \geq 0} -\gamma \left(\beta^2 \frac{1}{\beta s} + \frac{1}{s} - 2\beta \frac{1}{s}\right) - C_A^2 E_A \\
\iff &\max_{E_A \geq 0} -\gamma \left(\frac{1}{s}(1-\beta)\right) - C_A^2 E_A
\end{aligned}
$$

Alice's best response effort of the answer satisfies:

$$
\begin{aligned}
\frac{\gamma E_Q^2}{(E_Q E_A + E_Q s + E_A s)^2} &= C_A^2 \\
(E_Q E_A + E_Q s + E_A s)^2 &= \frac{\gamma E_Q^2 k_A^2}{\lambda_A^2} \\
E_A(E_Q + s) &= \frac{E_Q(\sqrt{\gamma} k_A - s\lambda_A)}{\lambda_A}
\end{aligned}
$$

The best response is then given by:

$$
R(E_Q) = \frac{E_Q(\sqrt{\gamma} k_A - s\lambda_A)}{\lambda_A(E_Q + s)}
$$

Since effort is bounded below at zero, it results that communication occurs if the cost of language is small enough and/or experience is high enough, such that $\sqrt{\gamma} k_A > s\lambda_A$.

3.2 Implications of the model after a variation in the cost of language

How effort decisions are affected by variations in the exogenous cost of language? Let λ'_A be the initial level of exogenous communication cost and λ''_A be the new level. The best response effort level of the answerer would then change by:

$$\Delta R(E_Q) = \frac{E_Q\left(\sqrt{\gamma}k_A - s\lambda''_A\right)}{\lambda''_A(E_Q+s)} - \frac{E_Q\left(\sqrt{\gamma}k_A - s\lambda'_A\right)}{\lambda'_A(E_Q+s)} \tag{1}$$

$$= \frac{E_Q\sqrt{\gamma}k_A(\lambda'_A - \lambda''_A)}{\lambda''_A\lambda'_A(E_Q+s)} = -\frac{E_Q\sqrt{\gamma}k_A\Delta\lambda_A}{\lambda''_A\lambda'_A(E_Q+s)}, \tag{2}$$

where $\Delta\lambda_A \equiv \lambda''_A - \lambda'_A$ is the size of the variation in the exogenous cost λ_A.

Equation 2 shows that, after a drop in the cost of language (i.e. $\Delta\lambda_A < 0$):

1. the effort choice of the answerer increases:

$$\Delta R(E_Q) > 0, \tag{3}$$

2. the change in the effort choice depends on the size of the change in the cost of language:

$$\frac{\partial \Delta R(E_Q)}{\partial \Delta\lambda_A} = -\frac{E_Q\sqrt{\gamma}k_A}{\lambda''_A\lambda'_A(E_Q+s)} > 0 \quad \text{if} \quad \Delta\lambda_A < 0 \tag{4}$$

3. the change in the effort is positive and concave on the effort made by the questioner:

$$\frac{\partial \Delta R(E_Q)}{\partial E_Q} = -\frac{\sqrt{\gamma}k_A\lambda''_A\lambda'_A\Delta\lambda_A s}{\left[\lambda''_A\lambda'_A(E_Q+s)\right]^2} > 0 \quad \text{if} \quad \Delta\lambda_A < 0 \tag{5}$$

$$\frac{\partial^2 \Delta R(E_Q)}{\partial E_Q^2} = \frac{2\sqrt{\gamma}k_A\Delta\lambda_A s}{\lambda''_A\lambda'_A(E_Q+s)^3} < 0 \quad \text{if} \quad \Delta\lambda_A < 0 \tag{6}$$

4. the change in the effort is positive and concave on the degree of incentive alignment:

$$\frac{\partial \Delta R(E_Q)}{\partial \gamma} = -\frac{E_Qk_A\Delta\lambda_A}{2\sqrt{\gamma}\lambda''_A\lambda'_A(E_Q+s)} > 0 \quad \text{if} \quad \Delta\lambda_A < 0 \tag{7}$$

$$\frac{\partial^2 \Delta R(E_Q)}{\partial \gamma_s^2} = \frac{E_Qk_A\Delta\lambda_A}{4\sqrt{\gamma}\gamma\lambda''_A\lambda'_A(E_Q+s)} < 0 \quad \text{if} \quad \Delta\lambda_A < 0 \tag{8}$$

3.3 Platform's trade off

The introduction of parallel versions of the platform in other languages allows a reduction in the cost of language for non-native English speakers. As shown, the model suggests that users benefiting from this lower cost will increase effort in answering.

Nevertheless, the platform may face other effects induced by the introduction of non-English versions of the website. In what follows, let λ''_A be the new level of the cost of language, and λ'_A the initial level, where $\lambda''_A < \lambda'_A$.

Potential increase in misleading answers

The model suggests that the effort choice is positive, i.e. it results in a published answer, if the following condition is satisfied:

$$\sqrt{\gamma} k_A > s\lambda_A.$$

In other words, the user provides an answer if the cost of language is sufficiently lower than her expertise. From the platform perspective anyway, a good answer is an answer that is both well written and accurate. The platform would then like that all participating users would have at least a minimum level of expertise, say $\bar{k}_A$.

A reduction in the cost of language may induce an increase in the number of answers by users that do not satisfy a minimum level of k_A. In fact, let $\hat{k}_A < \bar{k}_A$. Then, if:

$$\lambda''_A < \frac{\sqrt{\gamma}\hat{k}_A}{s} < \lambda'_A,$$

a user with an insufficient level of expertise would not answer questions on the English website, but she would provide answers in her native language website.

Ambiguous externalities on the English website

The introduction of non-English websites may induce both positive and negative effects on the English platform.

It is reasonable to assume that users are time-constrained and cannot just increase participation with no boundaries. If that is the case, users with a cost of language high enough would switch to their native-language website, and substitute effort from the English to the native language website.

If these switching users are high expertise users, the absence of their contributions in the English website may reduce the overall welfare in the English platform. On the contrary, if switching users are low expertise users, the English website may see an increase in the average quality of its content.

Inefficiency in knowledge aggregation versus fairness concerns

From an economic perspective, to have information shared in a multiplicity of languages is inefficient. To coordinate on the use of the same language would allow to maximize the aggregation of information and minimize search costs. Nevertheless, as noted by the Stack Overflow team itself, imposing a language over the others would exclude people who cannot learn that language, and would mean to arbitrarily decide what language should be the only one.[15]. This trade-off between efficiency and ethics is not only relevant for Stack Overflow, but in general on any discussion about centralization versus decentralization of languages (Ginsburgh and Weber 2011).

[15] https://stackoverflow.blog/2014/02/13/cant-we-all-be-reasonable-and-speak-english/

4 Data

To analyze communication strategies as a function of the exogenous cost of language, I retrieve the answers published in StackOverflow by two groups of users, the *Treatment* group and the *Control* group.

The *Treatment* group is composed of users who face a shock in the cost of language. In other words, this group includes users for whom English is not the native language, and who may incur a cost reduction once the website in their native language becomes available. I assume that users who published posts in a language different from English are native in that language.[16] This assumption implies that when a non-English website was released, users speaking that language were able to publish in their native language, facing a drop in the cost of communication. The date at which the website in the native language of a given user became available is defined as the treatment date for that user.

The selected sample is composed of all users who posted at least one answer in English before treatment and at least one answer in another language, i.e. Russian, Japanese, Portuguese, and Spanish.[17] Note that users could keep writing answers in English after the treatment, but this is not a condition to be in the sample.

The *Control* group instead is composed of a random sample of users who did not participate in any of the non-English platforms of Stack Overflow. I assume that these users were not affected by the introduction of StackOverflow platforms in languages different from English.

Table 1 reports the total number of answers and the number of users who wrote them contained in the sample used for the analysis. To identify the platforms, I use *SO* for the Stack Overflow in English, while I add the first letter of the language to *SO* for the other platform: *SOJ*, *SOP*,*SOR*, and *SOS* are, respectively, Stack Overflow in Japanese, Portuguese, Russian, and Spanish. The *Treatment status* indicates whether the authors published the answers before or after being treated. Figure 1 shows instead the sample size across time, with each platform's sample stacked vertically.

The data are right-censored, with the last date being the end of August 2017. Note that the authors in the Control group are randomly selected. On the contrary, the sample includes all answers of all users of the Treatment group.

4.1 Users' "adoption" of the native-language website

Once the platform implemented the non-English websites, treated users could participate using both their native language or English. Table 2 shows the extent to which treated users adopted the non-English websites. It reports the distribution of the number of answers published by each treated user in the sample before and after being treated. It is possible to notice that, on average, users kept writing in English even after their native

[16] To justify this assumption, note that English is the most common language used in the community of programmers, suggesting that if a person is fluent in English would just use the English website. This is confirmed by the fact that 99.8% of English-platform users contributing as well in other languages contribute in only one other language

[17] As of 2021, only those languages are available.

Group	Post in:	Treatment status	num. answers	num. authors	Earliest	Latest
Control	SO		6976	536	2008-09-16	2017-08-27
Treatment	SO	before	128984	2680	2008-08-12	2015-10-29
		after	100610	2089	2010-10-10	2017-08-28
	SOJ	after	3435	204	2014-10-10	2017-08-25
	SOP	after	30273	1183	2013-12-12	2017-08-27
	SOR	after	8448	137	2010-12-20	2017-08-28
	SOS	after	15139	1156	2015-10-30	2017-08-28

Table 1: Total number of answers in the sample, unique authors that wrote them, and dates of the earliest and latest answer in the group. Values are grouped by 1) Treatment group of the author (Treatment or Control), 2) platform, and 3) whether the author was treated at the time he/she wrote the answer.

language became available. On average, a treated user made 48 additional answers on the English website and 21 on the non-English website.

This statistic is anyway influenced by the amount of time the user in the sample has been treated. Figure 3 reports the same statistics but for only answers published within one year before and after treatment. This statistics are more comparable across languages, but the sample of authors used for the *before* treatment statistics may differ from the one used in the *after* statistics.

There is anyway substantial heterogeneity in behavior before and after the native-language website became available. Figure 2 shows how many treated users published a certain quantity of answers in the native-language website, conditional on how much they published in the English website before the native-language one became available.[18] It shows that users cluster on the extreme: either they produce a lot or very little in both websites, or they specifically prefer one of the two. Users relatively active in English publish very little in the native language, while users very little active in English increase a lot their participation in the native language. These differences suggest that some users have a low cost of using English, and the native-language platform did not bring many benefits to them, while other users have higher costs of using English, and, as a consequence, higher benefits from the introduction of the new websites.

4.2 Measure of answer's quality

A standard and simple measure of textual informativeness is the text length, measured, for instance, as the number of words used. This proxy for quality, as well as all alternatives that use text measures, is language-specific and not comparable across languages.

To overcome this issue, the paper proxies for quality using the number of separated pieces of code contained in the answer. More precisely, each answer is an *html*

[18] Sample of users active both in the year before and in the year after treatment

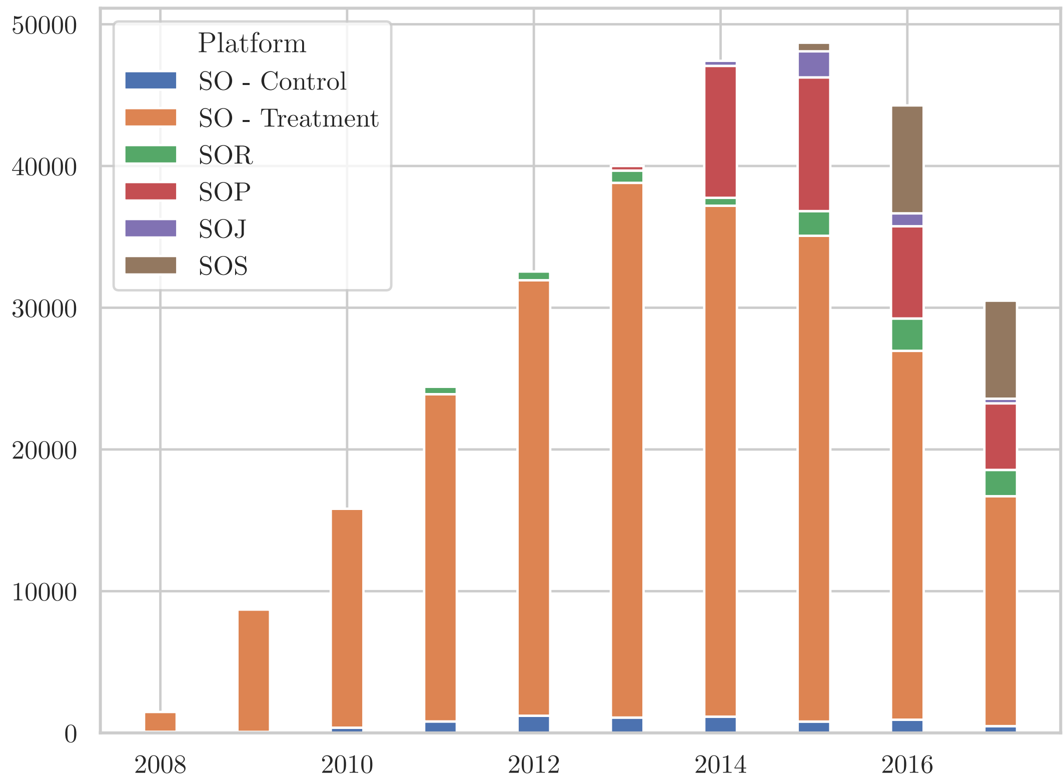


Figure 1: Number of answers in the sample for each year. The different colors identify the different platforms were those answers were published.

script. Once users include code snippets in the answer, they add *code* sections (i.e. <*code*>...< /*code*>) such that the code will appear in a separate box with a different color background. The box mimics a programming/statistical software's console and makes the code more readable. The proxy of quality is then defined as the number of *code* sections in the answer.

The intuition behind this measure is that a typical answer about programming would present some textual explanation and some code snippet to illustrate the solution. The presence of multiple snippets may indicate that either the answerer is providing several pieces of information, or that she is explaining one piece of information more clearly, with a step-by-step procedure. In both cases, more snippets suggest a higher quality of the answer.

Figure 3 shows that answers with more pieces of code on average obtain a higher number of points, where points are the result of up-votes and down-votes on the answer.

The pattern is consistent across websites, as it is possible to see in figure 4.

	SO (Before)	SO (After)	Not-SO (After)
count	2680.00	2089.00	2680.00
mean	48.13	48.16	21.38
std	184.75	224.15	134.70
min	1.00	1.00	1.00
25%	2.00	2.00	1.00
50%	5.00	7.00	2.00
75%	21.00	24.00	7.00
max	2848.00	4894.00	3759.00

Table 2: Statistics on the number of answers that each author published in English (Before and After the platform in her native language became available) and in her native language. Note: the row *count* identifies the number of users in the sample: this value is lower in the column "SO (After)" because some users switched fully to the non-English website.

4.3 Variables affecting communication effort

The model proposed explains communication effort as a function of the exogenous cost of language, the other-party communication effort, and the degree of incentive alignment. To capture these variables in the data, I identify several proxies.

Exogenous cost of language

To measure the degree of the exogenous cost of language that each user has, I exploit the rate at which users switch to their native language once allowed. The assumption underlying this proxy is that users with a higher cost of using English would be more willing to pay a switching cost and participate in their native language platform. More precisely this variable is measured as the number of posts (i.e. questions and answers) published in the native language over the total number of posts published after the native-language website became available.

The other-party communication effort

To proxy for the questioner's effort in communication, I adopt the number of pieces of code included in the question.

Incentive alignment

The data does not provide information on the degree to which the user internalizes the questioner's utility. Nevertheless, users can be incentivized by the questioner (or other participants) with reputation points. The platform allows the questioner to auction a certain amount of reputation points on a given question, and to promise to allocate these points to the user who would provide a satisfactory answer. The actioned points are allocated at the discretion of the questioner (even though some automatic allocation rules may apply in certain cases) and the questioner loose them even if the points are not allocated. This features allows for variation in virtual remuneration, which can proxy for the degree of incentive alignment between the communicating parties.

Empathy

	SO (Before)	SO (After)	Not-SO (After)
count	2074.00	1616.00	2480.00
mean	20.96	23.69	7.01
std	73.91	75.29	34.78
min	1.00	1.00	0.00
25%	1.00	2.00	0.00
50%	3.00	5.00	1.00
75%	10.00	14.00	3.00
max	1218.00	1204.00	943.00

Table 3: Statistics on the number of answers that each author published in English (Before and After the platform in her native language became available) and in her native language **within 1 year before and after the treatment**. Note: the row *count* identifies the number of users in the sample. **Given the two year time frame, some users may have contributed in that range before treatment but not after, or vice-versa**

Even if the communicating parties are not incentives aligned, the user providing the answer may put more effort if she feels empathetic with the author of the question. To capture the the degree of empathy, I use 1) whether the two communicating parties share the same language, 2) the type of profile picture displayed by the questioner, and 3) whether the questioner displays a full name (i.e. name and surname). All this information would allow the user answering to know whether the questioner shares the same nationality and group identity, which may affect her effort choice (Lyons 2017, BenYishay and Mobarak 2019, Ginsburgh and Weber 2020).

The variable that captures the commonality of language between the user and the questioner is a dummy equal to 1 if the questioner displays his location, and the language spoken in that location corresponds to the native language of the author of the answer. Note that this variable is based on the information available to the author of the answer, and it is not necessarily correct in reality. Nevertheless, to capture the degree of empathy, it is indeed relevant to rely only on the information available to the user. Note also that the "same language" variable takes always a value equal to 1 if the answer is published on a non-English website. The variable for the full name of the questioner is a dummy equal to 1 if the displayed name of the questioner matches the pattern of two words separated by a space and with capital letters. Finally, the type of questioner's picture corresponds to a categorical variable based on whether the questioner left the default avatar, added a personalized picture, or erased any picture.

Additional variables: competition in answering

One key limitation of the theoretical framework is that it abstracts from any strategic interaction between users external to the communicating parties. In reality, many users participate in the platform, and multiple users may provide answers for the same question. Communication effort may be affected by how many other users are potentially

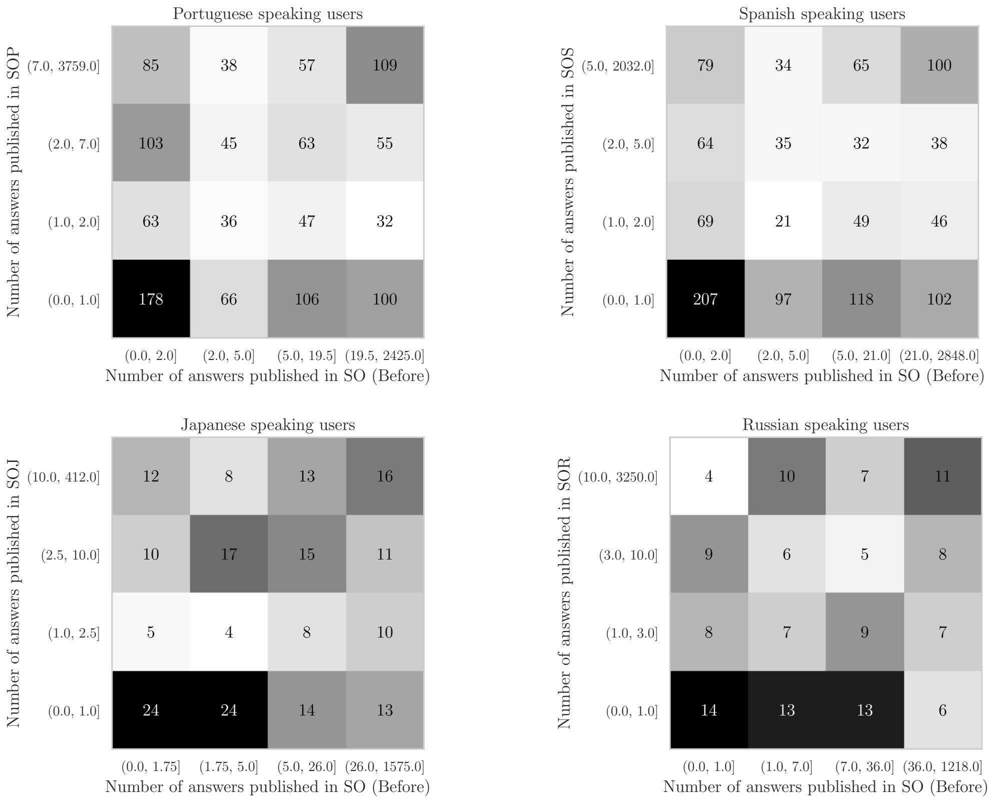


Figure 2: Distribution of users based on participation in English before the native-language platform became available and in the native language once it was available. Numbers in the plot correspond to the number of users in the sample who published a positive number of answers according to the respective intervals. Intervals are based on the 0.25, 0.5, and 0.75 quantiles of the respective distributions.

answering the same questions, and how many answers have been published already for the given question. To capture this form of "competition", I adopt two proxies: the total number of answers the question has received, and the total number of visits received by the question.

Table 4 reports descriptive statistics of the numeric variables, while the frequency distribution of the questioner's picture-types is displayed in figure 5.

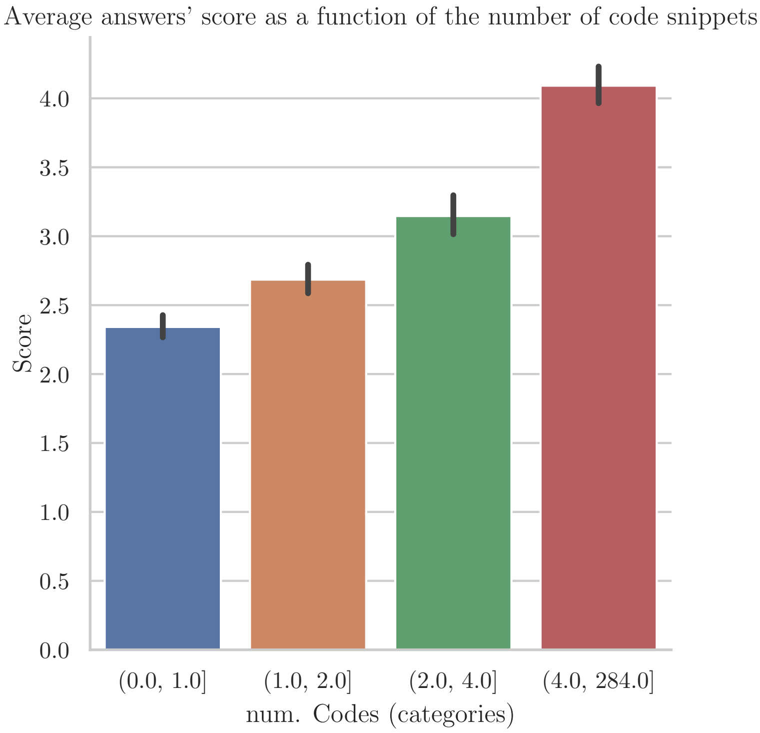


Figure 3: Average number of points obtained by answers, conditional on the answers having a certain number of pieces of code. Intervals of number of code snippets are based on the 0.25, 0.5 and 0.75 quantiles of the distribution across all answers in the sample. Vertical bars correspond to 95% confidence intervals.

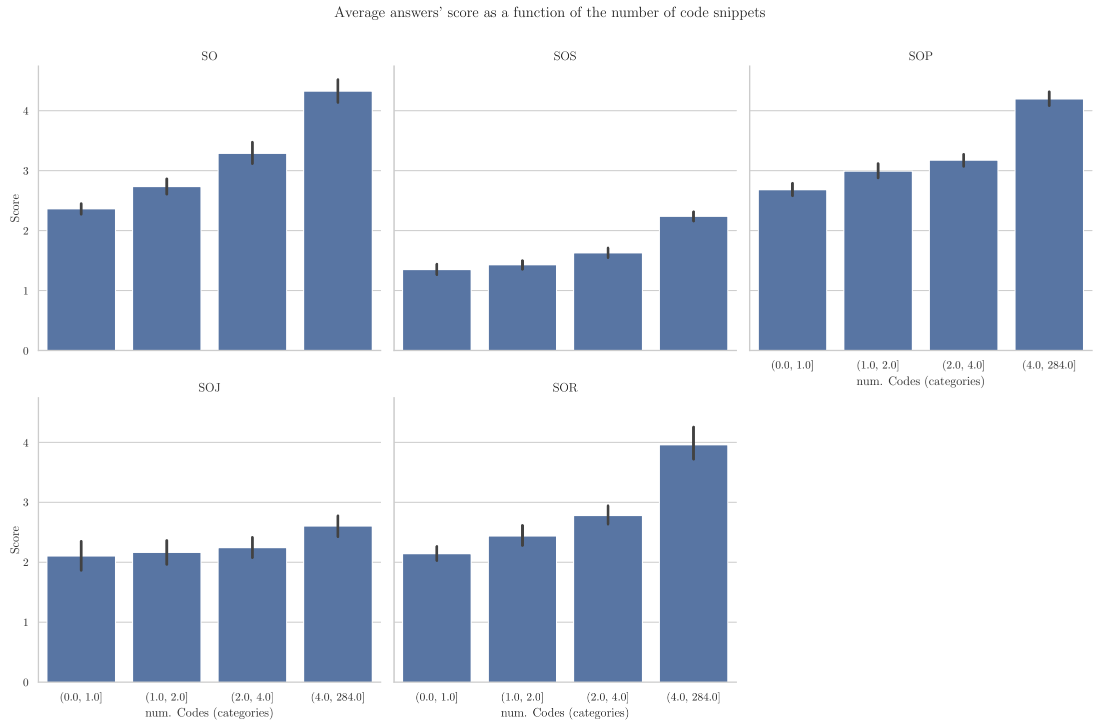


Figure 4: Average number of points obtained by answers, conditional on the answers having a certain number of pieces of code. Intervals of number of code snippets are based on the 0.25, 0.5 and 0.75 quantiles of the distribution across all answers in the sample. Vertical bars correspond to 95% confidence intervals.

	Same Language	Q. Full Name	# Answers	# Visits	Cost of English
Obs. Level	Answer level	Answer level	Answer level	Answer level	User level
count	286889	287319	293014	293014	2680
mean	0.21	0.25	2.77	8238.37	0.48
std	0.41	0.43	4.42	71859.1	0.36
min	0	0	0	5	0
25%	0	0	1	140	0.14
50%	0	0	2	584	0.42
75%	0	1	3	2290	0.88
max	1	1	518	8.67121e+06	1

Table 4: Descriptive statistics of variables affecting effort provision. Respectively, columns correspond to 1) a dummy equal to one if the author of the answer share the same native language as the questioner (Note that this variable is always one in the platforms using a language different from English); 2) a dummy equal to 1 if the questioner displays both name and surname; 3) the number of answers received by the question that the answer is answering to; 4) the number of visits received by the question that the answer is answering to; 5) user's cost level of using English, measured by the extent the user switched to her mother tongue when it became available.

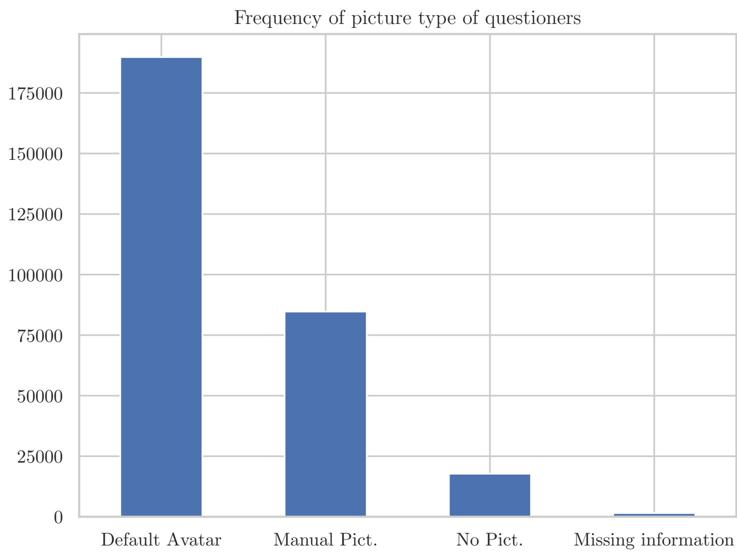


Figure 5: Frequency distribution across answers of the type of picture used by the author of the question which the answer is answering.

5 Descriptive illustration

Before moving to the analysis, a representation of the raw data may already provide suggestive evidence of users' behavior.

Figure 6 reports the average number of pieces of code, i.e. the measure of effort, made across answers before and after the non-English platforms became available. On the x-axis are reported 7-days periods before ad after. Note that they do not correspond to calendar weeks since the treatments are staggered. It is possible to see that while the average effort remains substantially similar on the English website, it is substantially higher in the non-English platforms. This shows that users on average include more pieces of code when they reply in their native language platform.

Figure 7 reports instead the average number of pieces of code made each calendar week included in the sample. Each graph plots the average effort of the Control group and of the users native in one of the non-English languages. Note that the data about English answers written by treated users include only the contributions of users native in that specific language. It is possible to see that in the raw data only Portuguese and Spanish speaking users include visibly more pieces of code in the native-language answers, suggesting that those groups of users may be particularly important for the final average effect.

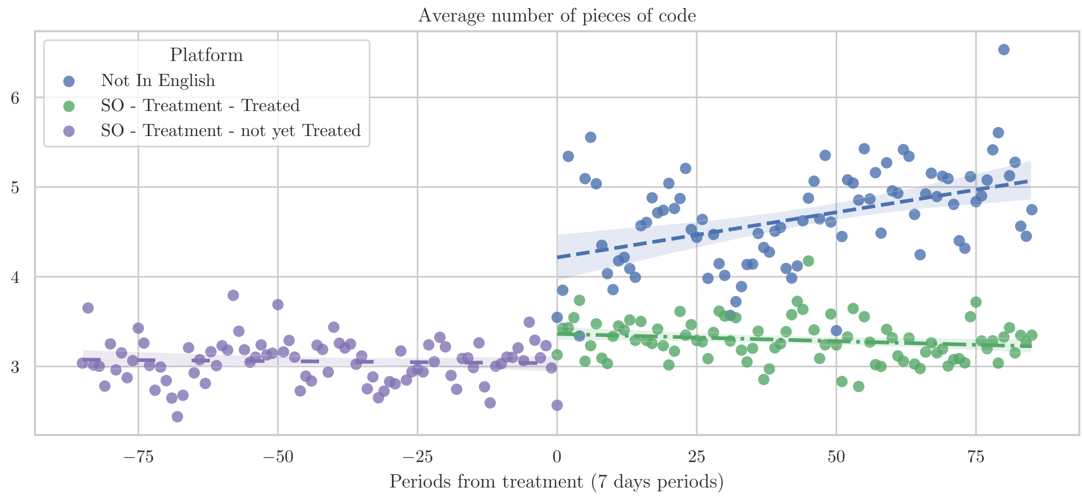


Figure 6: Average number of pieces of code in the 7-days periods before and after treatment, i.e. before and after the introduction of the native-language websites. In the after-period, separate colors discriminate between answers published in English and answers published in other languages.

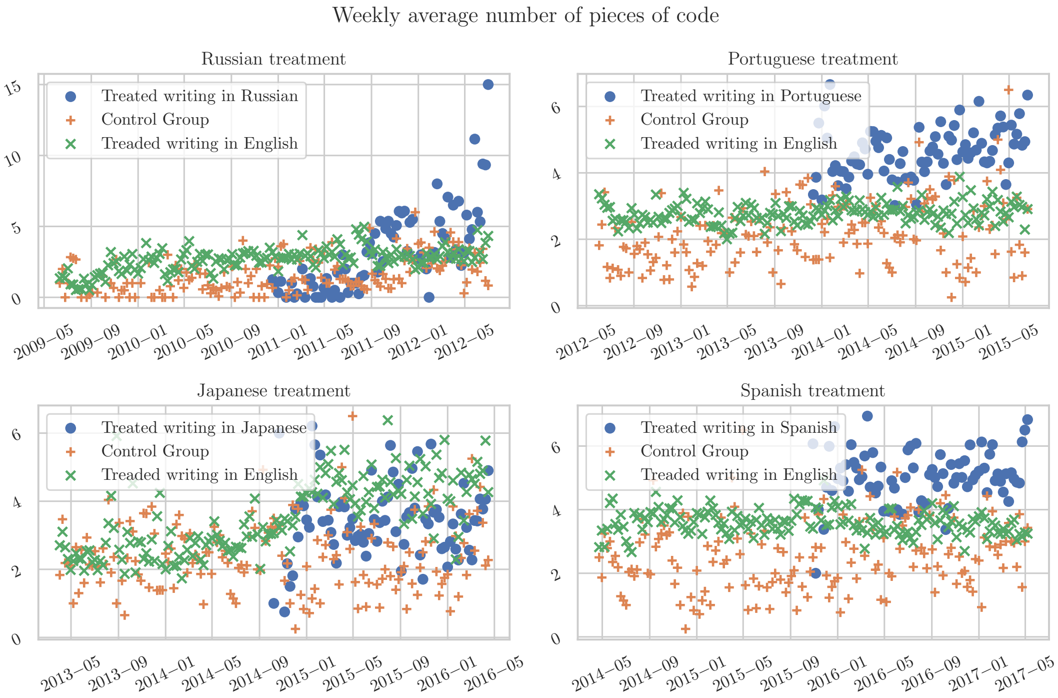


Figure 7: Average number of pieces of code across time. Graphs report data from the Control group of users (which is the same across graphs) and data of the treated users for each specific language, i.e. the sample of treated users is limited to those native in that specific language.

6 Empirical Analysis

The main interest of the paper consists in identifying the effect of a change in the exogenous cost of language on the communication effort choice. The implementation of Stack Overflow websites in languages different from English induced a reduction in the exogenous cost of language for users speaking those languages.

Let me define as *treated* units the non-native English users who have the possibility to participate in a Stack Overflow website using their native language.

The estimand of interest is then the average difference between the effort provided by the treated users when they are able to use their native language, and their effort in a potential scenario where they can only use English. More precisely, let j index the users, and t index the time period. The estimation target is:

$$\tau = \sum_{jt} w_{it}\tau_{jt}$$

$$\text{with} \quad \tau_{jt} = Y_{jt} - Y_{jt}(0)$$

$$\text{s.t.} \quad \text{j treated at time t}$$

Where w_{it} are non-stochastic weights, Y is the outcome variable, and $Y(0)$ is the potential outcome of a treated user if she would not be treated.

The identify this effect, I exploit the staggered implementations of the non-English websites, which allows me to compare 1) the treated units with units not yet treated, and 2) the treated units with units that will never be treated. To account for the individual and time fixed effects, the literature has traditionally adopted the so-called Two-Way Fixed Effect estimation method (TWFE) which consist of a linear regression of the outcome variable on individual fixed effects, time fixed effects, and a dummy equal to 1 when the unit is treated. The regression is estimated via OLS. Nevertheless, this approach has been proven to provide biased estimates in certain environments (Callaway and Sant'Anna (2020), de Chaisemartin and D'Haultfœuille (2020), Sun and Abraham (2020), Borusyak et al. (2021)). de Chaisemartin and D'Haultfœuille (2020) and (Callaway and Sant'Anna (2020) suggest alternative more robust approaches which identify the treatment effect relying only on the data just before and after the treatment of each cohort (i.e. the set of individuals treated at the same time. Differently, Borusyak et al. (2021) provides a different solution, based on the prediction of the unobserved potential outcome using a model *trained* on the "control" data. More precisely, the estimation strategy proceeds in three steps. First, it estimates via OLS the individual and time fixed-effects, using only non-treated observations, i.e. both not-yet treated and never treated units. It then predicts the potential counterfactual outcome $\tilde{Y}_{jt}(0)$ for the treated observations exploiting the estimates made in step one. This allows to compute the estimate of the treatment effect $\hat{\tau}_{jt} = Y_{jt} - \tilde{Y}_{jt}(0)$ for each observation. Finally, the third step averages the difference between observed and predicted outcomes across all observations with potentially heterogeneous weights across observations.

This estimation strategy relies on the parallel trend assumption, homoskedastic errors, and no-anticipation of the treatment. Note that in the context of this paper, even if the treatment is anticipated users do not incur in variation in the exogenous cost until they are treated, making the no-anticipation assumption naturally satisfied.

The method proposed by Borusyak et al. (2021) is the most appropriate given the data for two main reasons. The first is that it allows, differently from the other methods, to exploit the long time series of data available, while the second is that the data available is an unbalanced panel with heterogeneous gaps across individuals: the researcher does not observe behavior in the first period(s) after treatment for all users, as some users contribute rarely, and their first contribution in the native language may occur much later the creation of the new website. To focus only on the subset of data around the treatment date would then exclude some users from the sample, potentially creating selection issues.

6.1 Estimation

As a matter of comparability with traditional approaches, the analysis will provide estimation results for both the Two-Way Fixed Effects method (TWFE hereafter) and the method proposed by Borusyak et al. (2021) (BJS hereafter).

Let i index answers ,j index users (i.e. answers' authors), and t index time (weeks).

TWFE

The Two-Way Fixed Effect estimation approach would then estimate the treatment effect by estimation via Ordinary Least Squares of the following regression:

$$numCodes_{i(jt)} = \alpha_j + \alpha_t + \beta D_{jt} + \boldsymbol{W}'_{i(jt)}\gamma + \varepsilon_{i(jt)},$$

where $D_{i(jt)}$ is a dummy equal to 1 if author j at time t is able to use the website in her native language different from English, and β is the coefficient of interest, capturing the treatment effect. $\boldsymbol{W}_{i(jt)}$ is a vector of answer-specific control variables.

BJS

The alternative method proposed by Borusyak et al. (2021) instead estimates the treatment effect via a three-step procedure. First, it estimates via OLS a linear model on the non-treated sample:

$$\text{[Step 1]} \quad numCodes_{i(jt)} = \alpha_j + \alpha_t + \boldsymbol{W}'_{i(jt)}\gamma + \varepsilon_{i(jt)} \quad \text{if } j \text{ not treated at time } t,$$

then, it predicts, using the estimated model, the potential outcome of treated units if were untreated, and compute the observation-specific treatment effect:

$$\begin{aligned} \text{[Step 2]} \quad num\hat{C}odes_{i(jt)} &= \hat{\alpha}_j + \hat{\alpha}_t + \boldsymbol{W}'_{i(jt)}\hat{\gamma} \quad \text{if } j \text{ treated at time } t, \\ \hat{\tau}_{i(jt)} &= numCodes_{i(jt)} - num\hat{C}odes_{i(jt)} \quad \text{if } j \text{ treated at time } t. \end{aligned}$$

Finally, it averages all treatment effects, to obtain the average treatment effect:

$$\text{[Step 3]} \quad \hat{\tau} = \frac{1}{N} \sum_{i(jt)|\text{j treated at time t}} \hat{\tau}_{i(jt)}.$$

Table 5 reports the estimated treatment effect (i.e. *after*), corresponding to $\hat{\beta}$ in the Two-Way Fixed Effects specification and to $\hat{\tau}$ in the BJS specification. It shows that when users have the possibility to write answers in their native language, on average they include significantly more pieces of code. This result confirms the theoretical implication stated in equation 3, stating that a reduction in the exogenous cost of language induces an increase in communication effort.

6.1.1 Heterogeneity on the variation of the exogenous cost

The benefit from the reduction in the cost of language can be larger for individuals that have higher cost of using English. If we assume that all users face a similar language cost when they use their native language, users with higher cost of using English will face a larger drop in language cost once their native language gets available. This difference in the size of the cost reduction may result in a larger increase in effort provision, as stated in equation 4.

To address this heterogeneity, I categorize users by the share of posts (i.e. questions or answers) published in a non-English website relative to the total amount of posts published after the native-language website became available. The intuition is that

	(1) TWFE	(2) TWFE 1	(3) TWFE 2	(4) TWFE 3	(5) BJS	(6) BJS 1	(7) BJS 2	(8) BJS 3
after	0.392*	0.387*	0.388*	0.205*	0.656***	0.677***	0.683***	1.608***
	(0.107)	(0.111)	(0.111)	(0.0548)	(0.155)	(0.155)	(0.156)	(0.113)
Observations	293865	293007	293007	280495	293777	292846	292846	279649
pre_F					1.955	2.071	2.020	5.235
cse	2nd-lang	2nd-lang	2nd-lang	2nd-lang	User	User	User	User
Controls								
QEffort	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Competition	No	No	Yes	Yes	No	No	Yes	Yes
Empathy	No	No	No	Yes	No	No	No	Yes

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 5: Baseline Regressions' estimates where the dependent variable is the number of pieces of code. The estimate *after* corresponds to the average treatment effect, and corresponds to the parameters $\hat{\beta}$ or $\hat{\tau}$ if the specification adopted is the TWFE or the BJS respectively. *cse* represents the level at which the standard errors have been clustered: either at the users' native language (i.e. the level at which the treatment takes place) or at the user level.

assuming some switching costs from moving to a non-English website, the degree of switching would be proportional to the cost users incur in using English.

The measure adopted is based on users' endogenous choice of "adopting" the native-language platform, and cannot be considered as an exogenous characteristic. Nevertheless, it allows us to understand to what extent the cost of language is effectively driving the results. A larger measure corresponds to a more important switch to the native-language platform, and it would then give more weight to non-English answers, which are the ones on which we expect to see the difference in quality.

To test this prediction, I categorize the proxy for the cost of using English in four categories, as displayed in table 6. The boundaries of each category is based on the 25^{th}, 50^{th}, and 75^{th} quantiles of the distribution, as displayed in table 4. I then estimate separate treatment effects conditional on the cost of English level. More precisely, with c indexing the level of language cost, in the TWFE method the specification is the following:

$$numCodes_{i(jt)} = \alpha_j + \alpha_t + \sum_c \beta_c D_{jt} \mathbf{1}_{c(j)} + \boldsymbol{W}'_{i(jt)}\gamma + \varepsilon_{i(jt)},$$

where $\mathbf{1}_{c(j)}$ is an indicator function taking value 1 if the user j belongs to the level category c.

For what concern instead the BJS method, the cost-based estimates will be obtained by averaging the treatment effects within each category of cost:

$$\hat{\tau}_c = \frac{1}{N_c} \sum_{i(jt)|\text{j treated at time t}} \hat{\tau}_{i(jt)} \mathbf{1}_{c(j)}$$

where N_c is the number of answers in the sample written by treated users with a cost of English within the category c, .

Table 7 reports estimates for the four categories of exogenous cost of using English. Parameters corresponds to $\{\hat{\beta}_c\}_{\forall c}$ for the TWFE columns, and to $\{\hat{\tau}_c\}_{\forall c}$ for the BJS columns, where $c \in \{$Low, MediumLow, MediumHigh, High$\}$.

Estimates confirm the intuition stated in equation 4 and show that the larger the reduction in the cost of effort, the stronger the increase in communication effort.

	share of answers not in English in the after-period
Low	[0,0.14)
MediumLow	[0.14,0.42)
MediumHigh	[0.42,0.88)
High	[0.88,1]

Table 6: Categories for the exogenous cost of using English (boundaries rounded at 2 decimals)

	(1) TWFE	(2) TWFE 2	(3) BJS	(4) BJS 2
Low × after	0.0992	0.125	0.228	1.108***
	(0.114)	(0.102)	(0.182)	(0.114)
MediumLow × after	0.223	0.0878	0.474***	1.401***
	(0.122)	(0.106)	(0.141)	(0.128)
MediumHigh × after	0.659*	0.231	0.560***	1.695***
	(0.197)	(0.123)	(0.159)	(0.150)
High × after	1.476***	0.837*	1.883***	2.758***
	(0.142)	(0.175)	(0.271)	(0.261)
Observations	293007	280495	292846	279649
cse	2nd-lang	2nd-lang	User	User
Controls				
QEffort	Yes	Yes	Yes	Yes
Competition	Yes	Yes	Yes	Yes
Empathy	No	Yes	No	Yes

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 7: Estimates of average treatment effect by level of exogenous cost of using English. cse refers to clustered standard errors: TWFE model have standard errors clustered at the native language level (i.e. treatment level) while BJS at the user level.

6.1.2 The treatment effect depends on the questioner's effort

Equation 5 shows that the change in effort induced by the variation in exogenous cost increases on the effort made by the questioner. To test this prediction, I separately estimate the treatment effect by different levels of effort made by the questioners.

As a proxy for questioners' effort, I use the number of separated snippets of code that the questioner included in the question. Define this variable as $Qeffort$. I then bin this variable into four levels of effort, as described in table 8. The thresholds of each category correspond to the quartiles of variable $Qeffort$'s distribution.

The TWFE method's specification is then the following:

$$numCodes_{i(jt)} = \alpha_j + \alpha_t + \sum_{\eta} \beta_\eta D_{jt} \mathbf{1}_{\eta(j)} + \boldsymbol{W}'_{i(jt)}\gamma + \varepsilon_{i(jt)},$$

where η identifies the level of questioner's effort, and $\mathbf{1}_{\eta(i)}$ is an indicator function equal to 1 if the question that the answer is addressing contains a number of pieces of code within the η category.

For what concern instead the BJS method, average treatment effects are taken within each category:

$$\hat{\tau}_\eta = \frac{1}{N_\eta} \sum_{i(jt)|\text{j treated at time t}} \hat{\tau}_{i(jt)} \mathbf{1}_{\eta(j)}$$

where N_η is the number of answers in the sample written by treated users whose question is of quality level η.

Table 9 reports the estimate results: column 1 and 2 contain the $\{\hat{\beta}_\eta\}_{\forall\eta}$, while column 3 and 4 contain the $\{\hat{\tau}_\eta\}_{\forall\eta}$. Results confirm that the treatment effect grows with higher level of questioner's effort.

	average number of snippets of code in questions
Low	[0,1)
MediumLow	[1,2)
MediumHigh	[2,3)
High	[3,80]

Table 8: Categories for the effort level of the questioner

6.1.3 The treatment effect depends on the incentive alignment

According to the model, as shown in equation 7, the effect of a decrease in the exogenous cost of effort is proportional to the degree of incentive alignment between questioner and answerer.

To measure the incentive alignment between the two parties, I use the values of the so-called *bounties*. In Stack Overflow, any registered user who has more than 75 points can allocate some of her points to a question, and commit to giving these points to the author of an answer if judged satisfying enough[19]. Bounties can be considered virtual

[19]There are several rules in place to allocate the points to authors of answers. If the points are not allocated after a week, they are lost and do not return to the user who auctioned them.

	(1) TWFE	(2) TWFE 2	(3) BJS	(4) BJS 2
Low × after	0.143	-0.0524	0.374**	1.337***
	(0.129)	(0.0691)	(0.140)	(0.110)
MediumLow × after	0.581**	0.401**	0.868***	1.790***
	(0.0999)	(0.0541)	(0.164)	(0.118)
MediumHigh × after	0.579**	0.400**	0.884***	1.798***
	(0.103)	(0.0453)	(0.175)	(0.127)
High × after	0.592**	0.413***	0.977***	1.848***
	(0.0707)	(0.0229)	(0.187)	(0.141)
Observations	293007	280495	292846	279649
cse	2nd-lang	2nd-lang	User	User
Controls				
QEffort	Yes	Yes	Yes	Yes
Competition	Yes	Yes	Yes	Yes
Empathy	No	Yes	No	Yes

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 9: Estimates by level of questioner's effort. cse refers to clustered standard errors: TWFE model have standard errors clustered at the native language level (i.e. treatment level) while BJS at the user level.

payments and create a direct incentive to answer well the question. Note that if a bounty is active on a question, all other potential sources of points remain unchanged.

A user who answers a question with a bounty is then specifically more interested to make the questioner satisfied since the allocation of the bounty depends on that. I consider then this user more incentive aligned. The proxy for the incentive alignment of a user when he posts an answer is then given by the total sum of all bounties auctioned and not yet allocated before the answer is posted. I split this variable into four categories: the low category is composed just by the zero amount, while the other three categories are based on the 33^{rd} and 66^{th} quantiles of the distribution of the positive amounts. The categories are reported in table 10.

Similarly to previous heterogeneity analysis, the TWFE method estimates the treatment effects with the following specification:

$$numCodes_{i(jt)} = \alpha_j + \alpha_t + \sum_{\phi} \beta_\phi D_{jt} \mathbf{1}_{\phi(j)} + \boldsymbol{W}'_{i(jt)}\gamma + \varepsilon_{i(jt)},$$

where ϕ indexes the categories of the amount of active bounties open on the question addressed by the answer,, and $\mathbf{1}_{\phi(i)}$ is an indicator function equal to 1 if the question that the answer is addressing has an amount of ϕ active bounty points.

For what concern instead the BJS method, average treatment effects are taken within each category:

$$\hat{\tau}_\phi = \frac{1}{N_\phi} \sum_{i(jt)|\text{j treated at time t}} \hat{\tau}_{i(jt)} \mathbf{1}_{\eta(j)}$$

where N_ϕ is the number of answers in the sample written by treated users whose question has ϕ bounty points active.

Results are reported in table 11. They show that on average, the treatment effect is higher when authors are more incentive-aligned, confirming the prediction.

	amount of bounties
Low	0
MediumLow	50
MediumHigh	100
High	[150,1000]

Table 10: Categories for the amount of bounties allocated to questions that a user answered in a given week

	(1) TWFE	(2) TWFE 2	(3) BJS	(4) BJS 2
Low × after	0.373^{*}	0.190^{*}	0.666^{***}	1.594^{***}
	(0.110)	(0.0531)	(0.155)	(0.112)
MediumLow × after	1.235^{*}	1.045^{*}	1.645^{***}	2.310^{***}
	(0.287)	(0.237)	(0.253)	(0.238)
MediumHigh × after	2.272	2.112	2.759^{***}	3.642^{***}
	(0.818)	(0.860)	(0.370)	(0.346)
High × after	2.989^{***}	2.634^{**}	3.477^{***}	4.053^{***}
	(0.279)	(0.221)	(0.438)	(0.435)
Observations	293007	280495	292846	279649
cse			User	User
Controls				
QEffort	Yes	Yes	Yes	Yes
Competition	Yes	Yes	Yes	Yes
Empathy	No	Yes	No	Yes

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 11: Estimates by level of incentive alignment. cse refers to clustered standard errors: TWFE model have standard errors clustered at the native language level (i.e. treatment level) while BJS at the user level.

6.2 Robustness on quality measure

The number of pieces of code in the answer is a proxy for communication effort based on the characteristics of the message. Another approach to proxy for quality is based on observable outcomes, as, for instance, the questioner's appreciation of the answer.

StackOverflow allows authors of questions to "accept" an answer as "best answer". This action is not mandatory and does not depend on the number of answers provided to the same question. The action simply allows questioners to notify that the given answer solved or was good enough to solve the question they stated.

If authors of answers employ higher effort, the likelihood that their answer is accepted as the "best answer" should increase.

In this section, I estimate the treatment effect of a drop in the cost of language on the probability that the answer is accepted by the questioner as the best answer.

In a way similar to the previous analysis, I estimate the treatment effect using both the TWFE and the BJS methods.

TWFE

For the Two-Way Fixed Effects approach, the specification adopted is the following:

$$\mathbf{1}_{(\text{i(jt) is accepted})} = \alpha_j + \alpha_t + \beta^{BA} D_{jt} + \boldsymbol{W}'_{i(jt)}\boldsymbol{\gamma} + \varepsilon_{i(jt)},$$

where $\mathbf{1}_{(\text{i(jt) is accepted})}$ is an indicator function that takes value equal to 1 if answer i is accepted as "best answer" and 0 otherwise.

BJS

For what concerns the BJS method, I follow again the three-step procedure:

$$\begin{aligned}
&[\text{Step 1}] \quad \mathbf{1}_{(\text{i(jt) is accepted})} = \alpha_j + \alpha_t + \boldsymbol{W}'_{i(jt)}\boldsymbol{\gamma} + \varepsilon_{i(jt)} \quad \text{if } j \text{ not treated at time } t,\\
&[\text{Step 2}] \quad \widehat{\mathbf{1}_{(\text{i(jt) is accepted})}} = \hat{\alpha}_j + \hat{\alpha}_t + \boldsymbol{W}'_{i(jt)}\hat{\boldsymbol{\gamma}} \quad \text{if } j \text{ treated at time } t,\\
&\qquad\qquad \hat{\tau}^{BA}_{i(jt)} = \mathbf{1}_{(\text{i(jt) is accepted})} - \widehat{\mathbf{1}_{(\text{i(jt) is accepted})}} \quad \text{if } j \text{ treated at time } t.\\
&[\text{Step 3}] \quad \hat{\tau}^{BA} = \frac{1}{N} \sum_{i(jt)|\text{j treated at time t}} \hat{\tau}^{BA}_{i(jt)}.
\end{aligned}$$

Table 12 reports the estimates results. It shows that on average users are significantly more likely to have answers accepted once they are able to access the website in their native language.

6.3 Externalities on the English website

The introduction of clone websites in languages different from English may create externalities on the English website. These side effects may be both positive and negative. In a context where users are cognitive and/or time-constrained, the introduction of a website in a certain language may induce users native in that language to substitute effort from the English to the native-language website. This would potentially cause degradation of quality in the English website. Nevertheless, if users with a high cost of using English produce lower quality content in English, their shift to the native-language website should increase the average quality of the English platform.

To measure the net side effect on the English website, I estimate the treatment effect of the introduction of non-English platforms on the average quality of English

	(1) TWFE	(2) TWFE 1	(3) TWFE 2	(4) TWFE 3	(5) BJS	(6) BJS 1	(7) BJS 2	(8) BJS 3
after	0.0211***	0.0209***	0.0204**	0.00878	0.105***	0.105***	0.0931***	5.99e-15**
	(0.00244)	(0.00240)	(0.00243)	(0.00439)	(0.0168)	(0.0169)	(0.0157)	(1.85e-15)
Observations	293865	293007	293007	280495	293777	292846	292846	279649
pre_F					1.302	1.370	1.367	.
cse	2nd-lang	2nd-lang	2nd-lang	2nd-lang	User	User	User	User
Controls								
QEffort	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Competition	No	No	Yes	Yes	No	No	Yes	Yes
Empathy	No	Yes	No	Yes	No	Yes	No	Yes

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 12: Treatment effects on the probability of having an answer accepted as "best answer". cse refers to clustered standard errors: TWFE model have standard errors clustered at the native language level (i.e. treatment level) while BJS at the user level.

answers alone. More precisely, the externality effects are estimated with the following specifications:

TWFE

$$numCodes_{i(jt)} = \alpha_j + \alpha_t + \beta^{ext} D_{jt}\mathbf{1}_{(i \text{ is in English})} + \boldsymbol{W}'_{i(jt)}\boldsymbol{\gamma} + \varepsilon_{i(jt)}$$

BJS

$$\hat{\tau}^{ext} = \frac{1}{N_{eng}} \sum_{i(jt)|\text{j treated at time t}} \hat{\tau}_{i(jt)}\mathbf{1}_{(i \text{ is in English})}$$

Note that the sample on which is measured the treatment effect is the same as in previous analyses. The treatment effect is measured by comparing the average quality of answers written in English by non-native English speakers, after their native language became available, with their English answers written before, together with the ones of the control group.

As shown in table 13, in this context the different methodologies provide different results. The preferred approach (i.e. BJS) suggests that there are no significant externalities on the English platform. A deeper investigation would be necessary to understand the reason for the difference between the methods, and why the effect becomes positive and significant using BJS when controlling for the *Empathy* variables.

	(1) TWFE	(2) TWFE 1	(3) TWFE 2	(4) TWFE 3	(5) BJS	(6) BJS 1	(7) BJS 2	(8) BJS 3
after × InSo	-0.393	-0.413*	-0.418*	0.0394	0.203	0.209	0.216	1.050***
	(0.150)	(0.129)	(0.129)	(0.0341)	(0.154)	(0.156)	(0.156)	(0.0893)
Observations	293865	293007	293007	280495	236495	235574	235574	222887
pre_F					1.955	2.071	2.020	5.235
cse	2nd-lang	2nd-lang	2nd-lang	2nd-lang	User	User	User	User
Controls								
QEffort	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Competition	No	No	Yes	Yes	No	No	Yes	Yes
Empathy	No	Yes	No	Yes	No	Yes	No	Yes

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 13: Estimates of treatment effect on English answers' quality. cse refers to clustered standard errors: TWFE model have standard errors clustered at the native language level (i.e. treatment level) while BJS at the user level.

7 Conclusion

All digital platforms that serve or aim to serve people in several countries face the question of what language the website should be available in. This question is particularly relevant for products with network externalities and where the size of the community of users is determinant for the success of the platform, like StackOverflow.

In this work I address the question of the extent to which the use of a foreign language may affect communication quality, and if it is then important for the platforms to design systems that decrease those costs.

To achieve that, I exploit the case of StackOverflow, a question-and-answer website that, after facing those same questions, decided to create new websites that would allow users to speak their native language. The staggered introduction of these different websites allows increasing the validity of the results, as relatively similar users, i.e. non-native English programmers, have been treated at different timings.

This paper shows that there are important benefits to the quality of users' content if the platform succeeds in reducing users' exogenous communication costs. These effects may be multiplicative as they depend on the quality of the question too. Indeed, users who ask questions may also increase their effort when moving to their native-language website. The paper shows as well that, while cognitive constraints play an important role in message informativeness, incentives still matter. The increase in effort after the communication cost reduction is indeed higher if the information provider has at stake a larger reward.

These results suggest that StackOverflow chose wisely in adding versions of its platform in different languages. A major confirmation in this direction is the result that users partially switching to their native-language website did not worsen, on average, the quality of their contribution to the English website. Nevertheless, as discussed theoretically, other unintended consequences may have occurred, which would need to be analyzed empirically. First of all, it would be necessary to assess the welfare implications of a reduction in the ability of StackOverflow to aggregate information. To what extent some information is now available in Portuguese or Spanish, but not in English? The marginal value of answers not in English may be much lower, given the smaller audience. To what extent the additional website created more duplicate information? It is inefficient that the same question receives an answer in multiple languages, as effort is not optimized. Finally, language costs may be beneficial in refraining low-experienced users to participate. Their reduction may potentially cause an increase in the amount of misleading information provided.

The cost and benefits of centralization versus decentralization of the communication medium extend much more broadly than digital platforms. Indeed the question easily relates to the economic benefits to share the same language, as in the United States, versus the cultural diversity of Europe.

References

Arrow, K. J. (1974). *The Limits of Organization.* W. W. Norton & Company, Inc. 2

Asher, N. and A. Lascarides (2013). Strategic conversation. *Semantics & Pragmatics 6*(2), 1–62. 2

Austen-Smith, D. and J. S. Banks (2000). Cheap talk and burned money. *Journal of Economic Theory 91*(1), 1–16. 2

Bao, P., B. Hecht, S. Carton, M. Quaderi, M. Horn, and D. Gergle (2012, May). Omnipedia: bridging the wikipedia language gap. *CHI '12: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1075–1084. 3

Battiston, D., J. Blanes I Vidal, and T. Kirchmaier (2021, March). Face-to-face communication in organizations. *Review of Economic Studies 88*(2), 574–609. 4

BenYishay, A. and A. M. Mobarak (2019, May). Social Learning and Incentives for Experimentation and Communication. *The Review of Economic Studies 86*(3), 976–1009. 15

Blume, A. (2018, May). Failure of common knowledge of language in common-interest communication games. *Games and Economic Behavior 109*, 132–155. 2

Blume, A. and O. Board (2013, March). Language barriers. *Econometrica 81*(2), 781–812. 2

Blume, A., D. V. DeJong, Y.-G. Kim, and G. B. Sprinkle (2001). Evolution of communication with partial common interest. *Games and Economic Behavior 37*, 79–120. 4

Bolton, P. and M. Dewatripont (1994, November). The firm as a communication network. *The Quarterly Journal of Economics 109*(4), 809–839. 2

Borusyak, K., X. Jaravel, and J. Spiess (2021). Revisiting event study designs: Robust and efficient estimation. *Working Paper*. 3, 22, 23

Callaway, B. and P. H. C. Sant'Anna (2020, December). Difference-in-Differences with multiple time periods. *Journal of Econometrics*. 3, 22

Calvó-Armengol, A., J. de Martí, and A. Prat (2015). Communication and influence. *Theoretical Economics 10*, 649–690. 6

Crawford, V. P. and J. Sobel (1982). Strategic information transmission. *Econometrica 50*(6), 1431–1451. 2

Crémer, J., L. Garicano, and A. Prat (2007). Language and the theory of the firm. *The Quarterly Journal of Economics 122*(1), 373–407. 2

de Chaisemartin, C. and X. D'Haultfœuille (2020, September). Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects. *American Economic Review 110*(9), 2964–2996. 3, 22

Dewatripont, M. and J. Tirole (2005). Modes of communication. *Journal of Political Economy 113*(6), 1217–1238. 2

Diamond, J. (1991). *The Third Chimpanzee, The Evolution and Future of the Human Animal*. Hutchinson Radius. 2

Dilmé, F. (2016, December). Strategic communication by moderates and extremists. *Working paper*. 2

Dilmé, F. (2018, January). Optimal languages. *Working Paper*. 2

Gambetta, D. (2011). *Codes of the Underworld: How Criminals Communicate*. Princeton University Press. 2

Ginsburgh, V. and S. Weber (2011, April). *How Many Languages Do We Need?: The Economics of Linguistic Diversity*. Princeton University Press. 2, 3, 10

Ginsburgh, V. and S. Weber (2020, June). The Economics of Language. *Journal of Economic Literature 58*(2), 348–404. 15

Lafky, J. and A. J. Wilson (2020, January). Experimenting with incentives for information transmission: Quantity versus quality. *Journal of Economic Behavior and Organization 169*, 314–331. 4

Lohmann, J. (2011, February). Do language barriers affect trade? *Economics Letters 110*(2), 159–162. 4

Lyons, E. (2017, July). Team Production in International Labor Markets: Experimental Evidence from the Field. *American Economic Journal: Applied Economics 9*(3), 70–104. 15

Marschak, J. and R. Radner (1972). *Economic theory of teams.* Yale University Press. 2

Melitz, J. (2008, May). Language and foreign trade. *European Economic Review 52*(4), 667–699. 4

Sobel, J. (2013). Giving and receiving advice. In D. Acemoglu, M. Arellano, and E. Dekel (Eds.), *Advances in Economics and Econometrics*, Econometric Society Monographs, Chapter 10. Cambridge University Press. 2

Spence, M. (1973). Job market signaling. *The Quarterly Journal of Economics 87*(3). 2

Sun, L. and S. Abraham (2020, December). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*. 3, 22