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

Jacopo Bregolin*
University of Liverpool

April 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 investigate the trade-off faced by knowledge platforms in implementing their website in multiple languages. 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 show that the quality of communication improves by more than 24% when writers use their first language, rather than English, and answers are 7% more likely to solve the questioner's problem, a 20% increase from baseline. In addition, the size of the effect increases when the sender is more incentivised and when the questioner's effort is higher. With the introduction of other languages, the community size increases, but the quality of the contribution of the new joiners is lower. Finally, information is more dispersed. These results show that the platform should adopt multiple languages to maximise the quality of the information collected. The benefits anyway disappear if the community of users benefiting from the new languages is small: in that case, to maximize efficiency the platform should prefer one language only.

JEL Codes: D82, D83, L17, L86, M21, M52, 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 participants at the University of Liverpool Management School internal seminar and Digital Economy Network workshop (Norwich) 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 study to what extent the use of a foreign language affects effort choices in communication, and whether incentives matter. I then compare the advantages and disadvantages of reducing the cost of language by decentralizing the language used. 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 (Crémer et al. 2007, 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 on topics related to computer programming, 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 in addition to English.

The paper shows that users increase their communication effort by 24% when speaking in their native language, and are 7% more likely to provide satisfactory information. The effect on communication effort jumps to 110% when users are highly incentivized, and to 34% if the questioner puts a lot of effort. While these results suggest that the use of multiple languages is beneficial for communication, the paper shows that there are several drawbacks. Introducing multiple languages allows more people to share information, but the new joiners are on average less expert and decrease information quality. In addition, information gets more dispersed and potentially inefficiently duplicated.

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

[1] The economic literature has theoretically investigated both of these constraints but generally focuses 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.

and quality of the information provided: the platform has then all incentives to reduce the barriers to participation and to the provision of effort in communication.[2]

Should the platform have a unique website in English, or should it implement several websites in different languages? The optimal strategy is not trivial. By allowing users to communicate in languages different from English, the platform reduces communication costs and segregates communities. As shown in a simple theoretical framework, a reduction in the communication cost may have opposite implications. On one side, non-native English users may be able to increase information quality, if they were active in English, or start contributing if they were not. On the other side, new joiners may, on average, have lower expertise, so their additional contribution may decrease information quality. For what concerns the segregation of communities instead, on one side it is ethically preferable, as it does not impose a language over the others. On the other side anyway, the platform reduces its ability to aggregate information, causing an inefficient allocation of resources. This has been, for instance, found in Wikipedia, where some languages provide some information and others do not, and vice versa(Bao, Hecht, Carton, Quaderi, Horn, and Gergle 2012).

I can investigate this trade-off by observing users' behavior before and after their native language became available. Stack Overflow was created in 2008 in English, but, with time, the platform implemented additional websites in Russian, Portuguese, Japanese, and Spanish with the same purpose and function as the initial website. A unique Id for each user allows to track users across these websites.

I look at two measures of communication quality: one is directly addressing users' effort decision using message characteristics, while the second is a measure of communication outcome. More precisely, the former is based on the number of separate snippets of code included in the answer. Since questions relate to computer programming, a more developed and informative answer would include a step-by-step procedure that alternates text and code. More pieces of code would then signal higher quality.[3] The second measure instead exploits the fact that authors of the questions can *accept* one of the answers they receive if they consider it enough satisfactory. This choice is not mandatory, so it can reliably inform whether the questioner could solve his problem with the information received.

For what concerns the proxy for the degree of incentive alignment instead, I exploit another feature of the website called *bounties*. Stack Overflow users can auction reputation points (i.e. virtual rewards) on given questions. In other words, they can commit to providing a reward to the author of an answer considered enough satisfactory. The size of the number of points at stake could then provide a measure of how much the author of the answer is incentivised.

The empirical analysis for the effect of a drop in the cost of language uses a staggered difference-in-difference approach. In other words, I execute a regression analysis at 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]Note that I am not measuring the length of the code.

answer level with communication quality as the dependent variable, the availability of the user's native-language website as treatment dummy, and time and user fixed effects. 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.[4] I then proceed with a heterogeneous analysis by interacting the treatment dummy with different levels of 1) questioner's effort, 2) incentive alignment, and 3) the degree to which users started contributing in their native-language website. Finally, I evaluate for externalities on the English website by limiting the analysis to only English answers.

Other dimensions of the trade-off are analyzed more descriptively by comparing the non-native-English users who were contributing in English before their native language became available with those who were not. This comparison allows us to assess both differences in community size and quality of contributions. Finally, I measure the dispersion of information using *tags*, i.e. labels attached to questions to categorize their content. If the same tag appears across different languages, then some information may have been provided multiple times, meaning that some efficiency was lost. At the same time, if some topics are treated in some languages but not English, then the multiplicity of languages suggests that the platform lost some ability to aggregate information.

Overall the paper shows that the trade-off, which emerges from the theoretical framework, is confirmed in the data. The introduction of native-language websites increases communication effort by 24% and even higher if the user is incentivized or if the questioner puts more effort. At the same time, there are no clear negative externalities for the English website. In addition, at least 42% of non-native English speaking users are likely to not have joined the platform in absence of their native-language website, implying a substantial increase in community size.[5] Nevertheless, new contributors appear to provide significantly lower quality answers in their native language, compared to users already active in English before the native-language website became available. In addition, there is a substantial overlap of topics across languages: out of 247 topics, 83 appear in at least 2 languages. At the same time, 12 topics out of the 247 are not treated in English at all, meaning that the platform did not optimally aggregate information. These facts suggest that there is potential for efficiency gains by imposing a single language. From these results we can infer that a knowledge platform highly benefits from multiple languages, as they both increase the community size and the quality of information collected. Nevertheless, these benefits shade away if the non-native-English users are very few, or they have a very low cost of using English. In this case, from an efficiency standpoint, a single language (or a limited number of languages) is preferable.

While this trade-off is particularly relevant for knowledge platforms, it is generalizable to any economic environment or institutions where language diversity imposes the critical choice between centralization or decentralization of language. A typical example

[4] The approach by Borusyak et al. (2021) solves for econometric issues identified by the literature (Callaway and Sant'Anna 2020, de Chaisemartin and D'Haultfœuille 2020, Sun and Abraham 2020, Borusyak et al. 2021).

[5] It has been shown that larger community size is beneficial not only because they constitute a larger base of contributors, but because it creates additional incentives for contributing users to provide content (Zhang and Zhu 2011)

is national states (Ginsburgh and Weber 2011), where we saw the homogenization of language, like in France (Blanc and Kubo 2021), or the maintaining of language diversity, like in Spain. Another example is the firms choosing between common or specialized languages (Crémer et al. 2007). Finally, the trade-off is relevant in international trade, where a common language is necessary to find agreements, but language costs may prevent efficient interactions (Melitz 2008, Lohmann 2011).

To my knowledge, this is the first paper to empirically quantify the role 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). The works by Battiston, Blanes I Vidal, and Kirchmaier (2021) and Guillouët, Khandelwal, Macchiavello, and Teachout (2021) also study exogenous communication costs and their effect on performance. The former anyway focuses on communication frictions arising from not being able to talk face-to-face, rather than the language itself. The latter is more specifically looking at the exogenous costs of language differences and their effect on productivity. The paper exploits a field experiment and shows that a reduction in language costs improves productivity. This result reinforces what I find. Nevertheless, they do not observe the exact communication, but only communication outcomes, so they do not quantify changes in communication effort (apart from the frequency of interactions). In addition, their paper is silent on the role of the other party's effort, and the incentives of the information holder.

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 presents the analysis for the effect of the cost of language on communication effort, section 6 discusses the trade-off faced by the platform, 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 are that it is crowd-based and free of charge. In other words, any internet user who registers (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 from the language, their function is 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 rollout of the final version.[12]

Figure 1 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 A.1 provides more details on the introduction of the websites

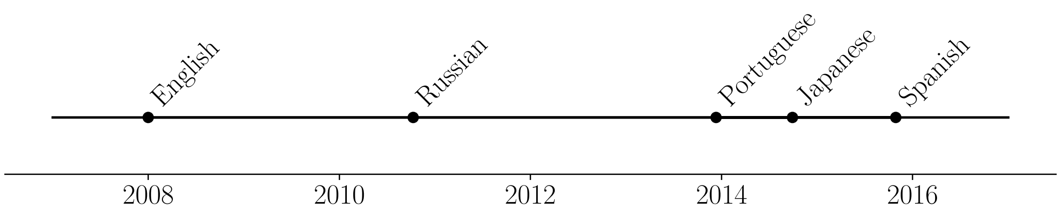

Figure 1: Timeline of the introduction of Stack Overflow websites.

HashCode were imported in the Russian version of Stack Overflow. Formally then, the Russian version of Stack Overflow appeared in 2015. Figure 1 reports the 2010 date as the data includes all the HashCode content.

3 Theoretical Illustration

In this section, I present a simple framework to provide a general overview of what type of factors affect communication effort decisions. Its aim is to guide the empirical analysis and understand what type of conditions affect the intensive and extensive margin of communication decisions.

The framework models unilateral information transmission in the Stack Overflow 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:

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

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, $\varepsilon \sim \mathcal{N}\left(0, \frac{1}{E_Q}\right)$ and $\eta \sim \mathcal{N}\left(0, \frac{1}{E_A}\right)$.

The answer displayed in the platform is then a realization of the message distribution, which Bob can observe. Finally, let the action $a \in (-\infty, \infty)$ 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 will use the observed message to update his beliefs about the true value of θ. Alice will internalize Bob's utility to a certain degree $\gamma \in [0,1]$. In addition, since she knows Bob's prior and the message realization, she can anticipate Bob's action, given the message.

3.1 Optimal effort strategies

For a given question, how much effort Alice will decide to make?[15] Proceeding backward, Bob selects the action a^* such that:

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

[15]Details on the steps are provided in appendix B

The optimal action is then given by:

$$a^* = \mathbb{E}[\theta|m] = \beta m \quad \text{with} \quad \beta \equiv \frac{E_Q E_A}{E_Q E_A + E_Q s + E_A s}$$

To find her optimal effort level, Alice solves:

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

Her 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 on the effort made by the questioner:

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

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

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

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 2 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.

[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:	Status	#answers	#authors	Earliest	Latest
Control	SO		6976	536	2008-09-16	2017-08-27
Treatment	SO	Not yet Treated	128984	2680	2008-08-12	2015-10-29
		Treated	100610	2089	2010-10-10	2017-08-28
	SOJ	Treated	3435	204	2014-10-10	2017-08-25
	SOP	Treated	30273	1183	2013-12-12	2017-08-27
	SOR	Treated	8448	137	2010-12-20	2017-08-28
	SOS	Treated	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.

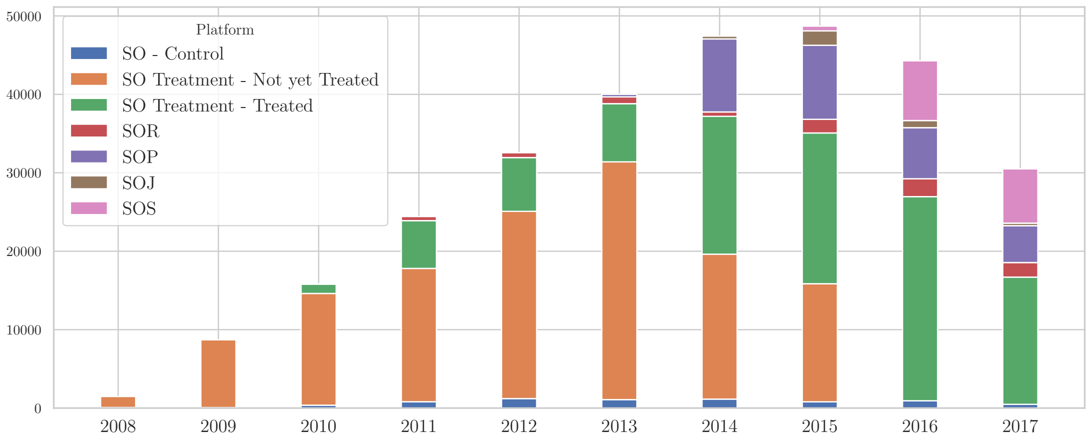

Figure 2: Number of answers in the sample for each year. The different colors identify the different platforms where those answers were published. SO, which corresponds to Stack Overflow in English, is split by answers published by users never treated (Control), users not yet treated, and users already treated

4.1 Users' "adoption" of the native-language website and externalities in the English 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 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. These 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 3 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. The figure reports separate plots for each website. The figure shows that users cluster on the extreme. On one side, as a general pattern across websites, some users who produce a lot in the English website before treatment also contribute a lot in the non-English one, and vice-versa. On the other side, some users contribute a lot in their native language after contributing little in English and vice-versa. The latter pattern is more clear in the Spanish and Portuguese websites. It suggests 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.

To quantify the extent to which users switch to their native language platform, I compute users' rate of contribution to the native language website over the total amount of contribution after treatment. For this specific purpose, and differently from previous statistics, I consider *contribution* both questions and answers. The measure is the result of the total number of questions and answers published by the user in the native-language website over the total number of questions and answers published after treatment (both in English and in the native language). Figure 4 shows the distribution of the measure, which confirms the above discussion: generally users, after treatment either contribute only to the native-language website, or they mainly contribute in the English one. (Note that, by the construction of the data, all users in the sample contribute to the native-language website, which explains the absence of a high spike at 0).

Figure 5 instead shows how many users contributed a certain amount of answers in English after being treated, and based on their contribution before treatment. It shows that a significant amount of users stop contributing to the English website. In general, anyway, those are users who were already contributing little before. The Spanish and Japanese users tend to either keep the same level of contribution or decrease it, while

	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.

Portuguese and Russian users tend to keep the same level or increase it. Russian users in particular seem to increase contributions in English. This anyway may be driven by Russian specific features, as discussed in section 2.2

	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**

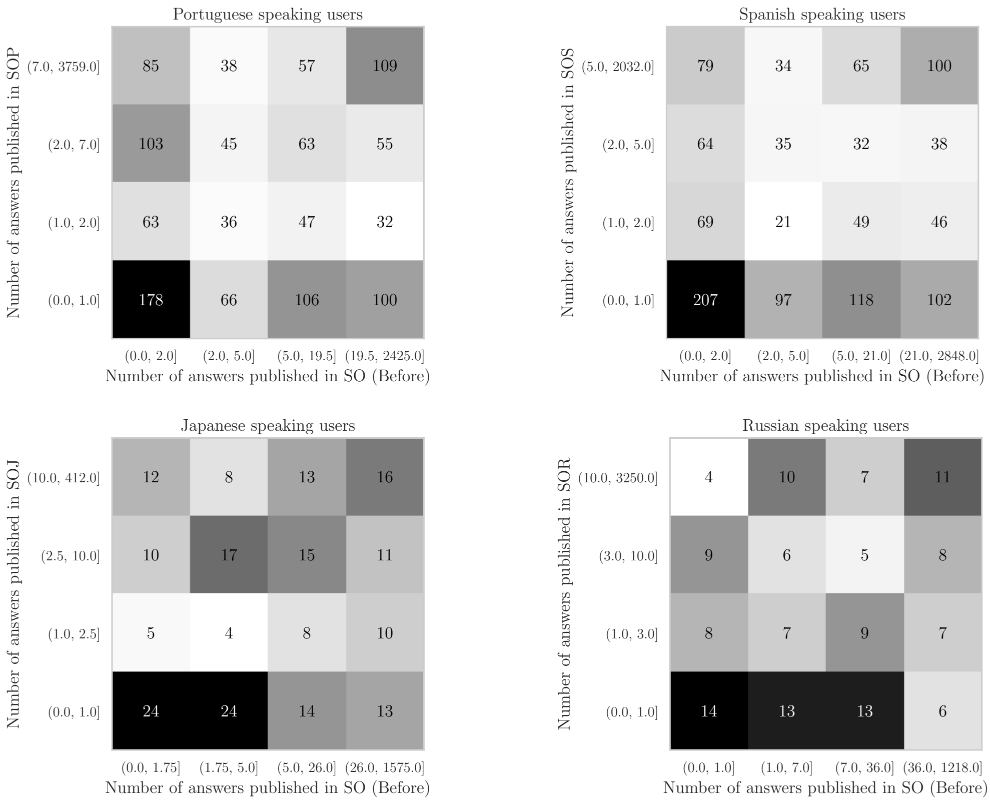

Figure 3: Distribution of users based on participation in English before the native-language platform became available and in the native language once it was available. Numbers in the plot correspond to the number of users in the sample who published a positive number of answers according to the respective intervals. Intervals are based on the 0.25, 0.5, and 0.75 quantiles of the respective distributions.

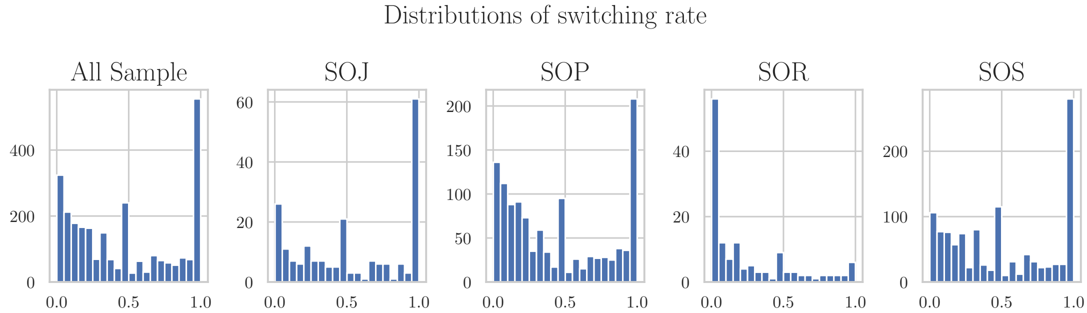

Figure 4: Distribution of switching rate measured as the total number of question and answers published by the user in the native-language website over the total number of questions and answers published after treatment.

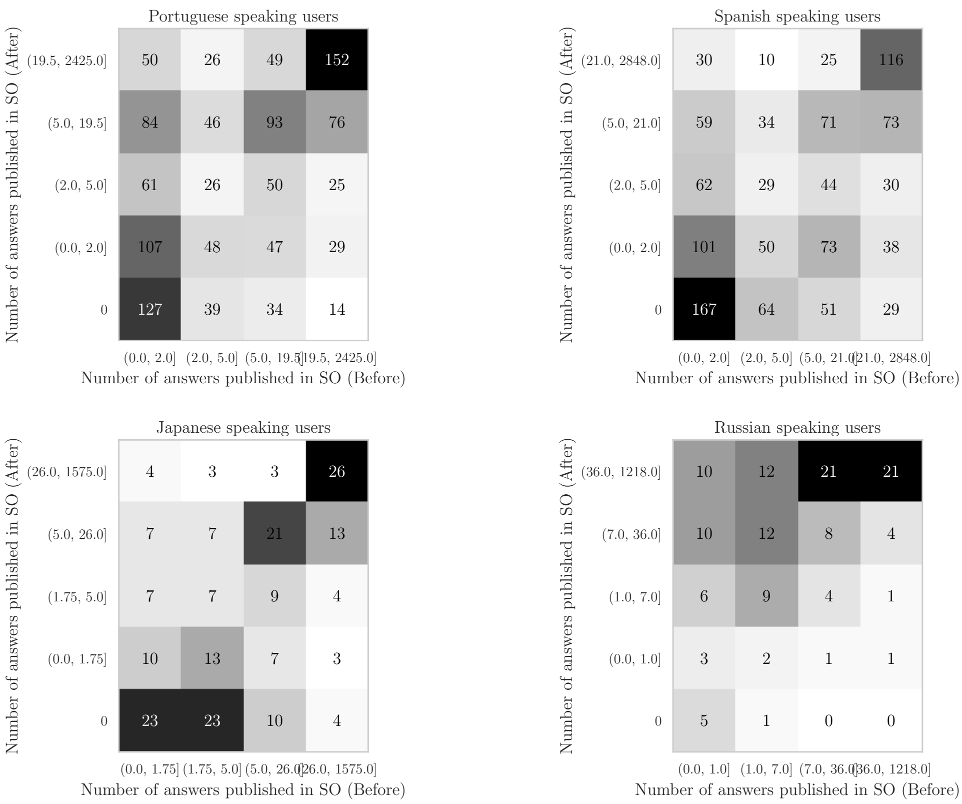

Figure 5: Distribution of users based on participation in English before and after treatment. Numbers in the plot correspond to the number of users in the sample who published 0 or more answers in English after treatment, based on their contribution before treatment. Intervals are based on the 0.25, 0.5, and 0.75 quantiles of the distributions of contributions before treatment.

4.2 Measure of answer 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* 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. In the appendix, figure 13 shows an example of an answer with two snippets of code. The intuition behind this measure is that a typical answer about programming would include 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 higher informativeness of the answer. Table 4 reports the distribution of the number of code snippets across answers.

Distribution Num. Codes	mean	std	min	25%	50%	75%	max
Full sample	3.37	4.76	0.0	1.0	2.0	4.0	284.0
Before Treatment - SO	2.71	3.99	0.0	1.0	2.0	3.0	284.0
After Treatment - SO	3.55	4.65	0.0	1.0	2.0	4.0	153.0
After Treatment - SOJ	3.70	4.65	0.0	1.0	2.0	5.0	52.0
After Treatment - SOP	4.60	6.38	0.0	1.0	3.0	6.0	186.0
After Treatment - SOR	4.49	6.18	0.0	1.0	3.0	6.0	120.0
After Treatment - SOS	5.00	5.93	0.0	1.0	3.0	6.0	129.0
Never treated	2.26	3.34	0.0	0.0	1.0	3.0	59.0

Table 4: Distribution of the number of pieces of Codes across all answers of the sample

Another possible way to proxy for answers' quality is to measure the degree of appreciation from the community. The data provides two possible indicators of such measures: answers acceptances and up-votes. The user who asked the question can *accept* one of the answers as *best answer*. This indicates that that answer was the one to solve his problem. At the same time, every registered user can up-vote (or down-vote) answers, similarly to how users allocate *likes* in other platforms.[18]. These measures anyway depend on time: users could accept or up-vote answers later than when the answer is published. This means that, as a measure of quality, they are not comparable across

[18] There are some exceptions on who can vote content. For details, see https://stackoverflow.com/help/privileges/vote-up

answers recently published and answers on the website for a long time. Nevertheless, both measures correlate with the number of pieces of code. Figure 6 shows that answers with more pieces of code on average obtain a higher score, where the score is the number of the up-votes net of downvotes that the answer received. It also shows that *accepted* answers include, on average, a higher number of pieces of code. The pattern is consistent across websites, as it is possible to see in figure 7, but in general, does not apply to answers with zero pieces of code. This suggests that there may be two different types of questions, one that requires some code in the answer and one that does not. I will address this issue in the analysis via robustness checks.

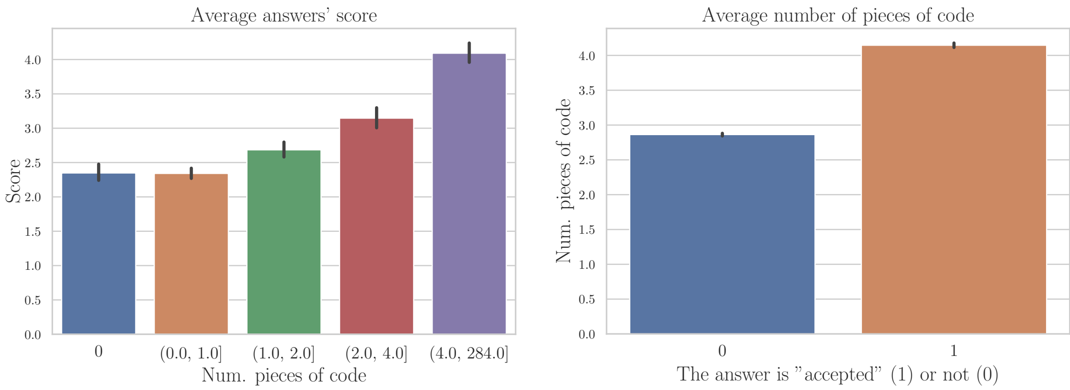

Figure 6: [**Left**]Average score obtained by answers, conditional on the answers having a certain number of pieces of code. Intervals of number of code snippets are based on the 0.25, 0.5 and 0.75 quantiles of the distribution across all answers in the sample. Vertical bars correspond to 95% confidence intervals.[**Right**] Average number of pieces of code across non-accepted (0) or accepted (1) answers. Vertical bars are 95% confidence intervals via bootstrapping

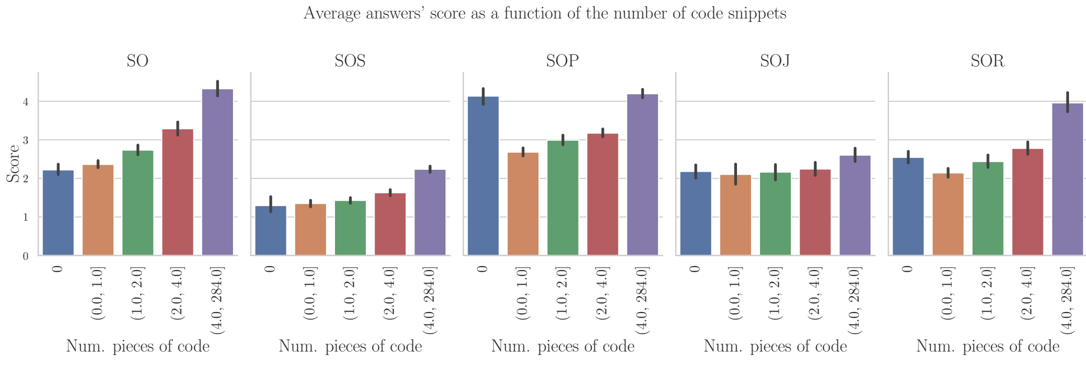

Figure 7: Average number of points obtained by answers, conditional on the answers having a certain number of pieces of code. Intervals of number of code snippets are based on the 0.25, 0.5 and 0.75 quantiles of the distribution across all answers in the sample. Vertical bars correspond to 95% confidence intervals. Vertical bars are 95% confidence intervals via bootstrapping

4.3 Variables affecting communication effort potentially correlated with the treatment dummy

The quality of the question
As suggested by the theoretical framework presented in section 3, under communication complementarities the sender chooses a certain level of effort as a function of the receiver's effort. If the author of the question increases his effort when writing on the non-English website, then the identification may be harmed. Since I observe the questions, I proxy for the questioner's effort in communication with the number of pieces of code included in the question, that is, the same measure of quality used for also for the answers.

Incentive alignment
Similarly, the theoretical framework suggests that communication effort depends also on the degree of 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 loses them even if the points are not allocated. This feature allows for variation in virtual remuneration, which can proxy for the degree of incentive alignment between the communicating parties.

Empathy
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 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.[19] 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

[19] Since I do not have reliable ways to identify the nationality of users in the control group, this variable is missing for those users.

corresponds to a categorical variable based on whether the questioner left the default avatar, added a personalized picture, or erased any picture.

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 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 5 reports descriptive statistics of the numeric variables, while the frequency distribution of the questioner's picture-types is displayed in figure 8.

	Same Language	Q. Full Name	# Answers	# Views
mean	0.21	0.25	2.77	8238.37
std	0.41	0.43	4.42	71859.14
min	0.00	0.00	0.00	5.00
25%	0.00	0.00	1.00	140.00
50%	0.00	0.00	2.00	584.00
75%	0.00	1.00	3.00	2290.00
max	1.00	1.00	518.00	8671208.00
Sample size	286889.00	287319.00	293014.00	293014.00

Table 5: 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.

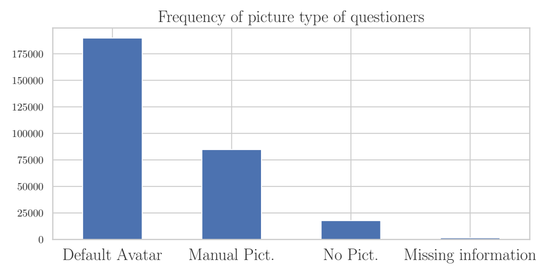

Figure 8: Frequency distribution across answers of the type of picture used by the author of the question which the answer is answering.

5 The effect of the cost of language on communication effort

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

Figure 9 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 10 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.

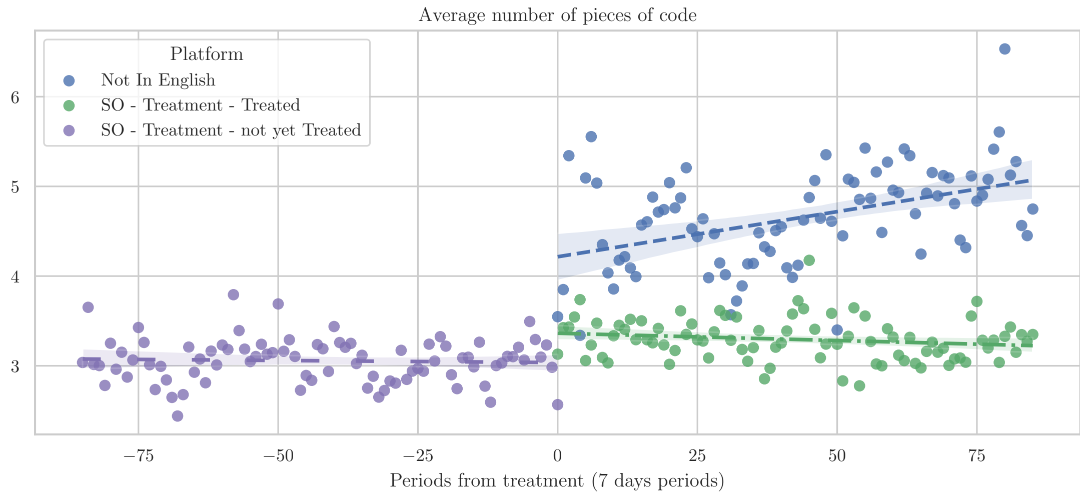

Figure 9: Average number of pieces of code in the 7-days periods before and after treatment, i.e. before and after the introduction of the native-language websites. In the after-period, separate colors discriminate between answers published in English and answers published in other languages.

Weekly average number of pieces of code

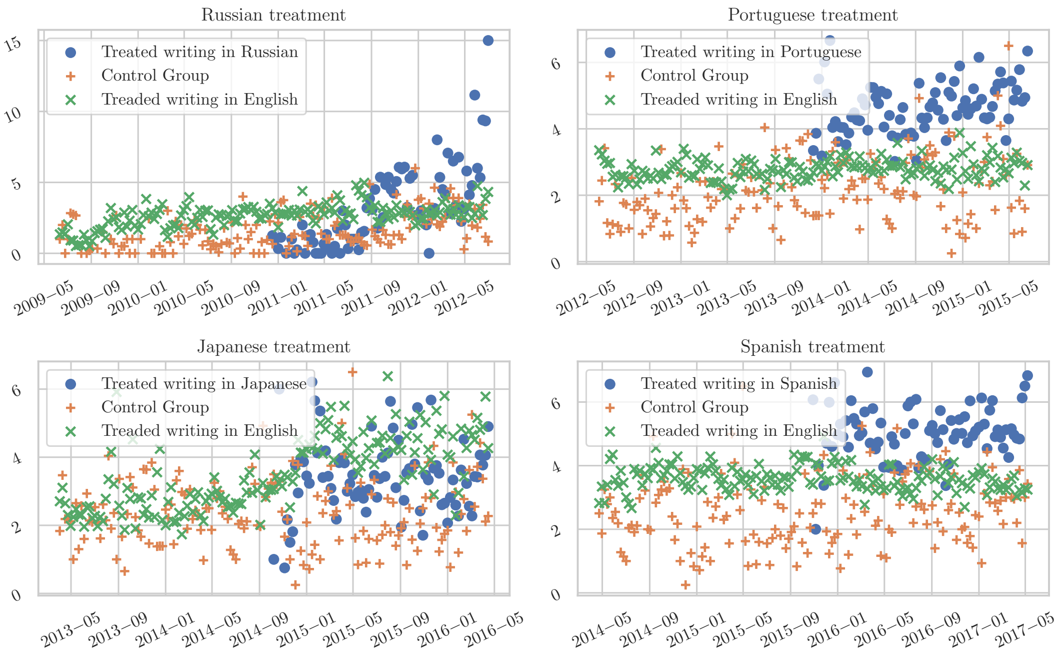

Figure 10: Average number of pieces of code across time. Graphs report data from the Control group of users (which is the same across graphs) and data of the treated users for each specific language, i.e. the sample of treated users is limited to those native in that specific language.

5.1 Estimation

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 can 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 can 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.

To 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, in the context of the data used for the analysis, this approach is likely to provide biased estimates.[20] To overcome this issue, I use the estimation strategy proposed by Borusyak et al. (2021), which is based on the prediction of the unobserved potential outcome using a model *trained* on the "control" data.[21] 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.

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:

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

[20] As discussed by several papers (Callaway and Sant'Anna 2020, de Chaisemartin and D'Haultfœuille 2020, Sun and Abraham 2020, Goodman-Bacon 2021, Borusyak et al. 2021) the two-way fixed effect estimation procedure estimates the treatment effect as a weighted average of treatment effects for each user × period cell. The weights sum to one, but may be negative. This fact may be an issue in the context of this paper. I cannot rule out that the treatment effect is heterogeneous across time and users. Users may potentially take time to adjust to the new environment, and certainly users with higher cost of using English may be more impacted. This may cause biased estimates using OLS, potentially even of opposite sign if the treatment effect increases over time.

[21] The literature has proposed other solutions, e.g. de Chaisemartin and D'Haultfœuille (2020) and Callaway and Sant'Anna (2020) suggest alternatives that rely only on the data just before and after the treatment of each cohort (i.e. the set of individuals treated at the same time). In the context of this paper, those solutions are less preferable because I observe an unbalanced panel (not all users participate every week). The selection of data may cause the creation of biased comparison groups.

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 6 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 can 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.[22]

5.1.1 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 7. The thresholds of each category correspond to the quartiles of variable $Qeffort$'s distribution.

The TWFE method's specification is then the following:

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

[22] As discussed in the data section, the correlation between the number of pieces of code and up-votes is not satisfied when the answers have zero pieces of code. This may suggest that some answers do not need to include any code. table 18 in the appendix reports the regression results after dropping all answers with zero pieces of code and selecting users that, given the remaining answers, were active both before and after treatment.

	(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***	0.663***
	(0.107)	(0.111)	(0.111)	(0.0551)	(0.155)	(0.155)	(0.156)	(0.169)
Observations	293777	292919	292919	280407	293777	292846	292846	199564
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 6: 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.

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 8 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	(3,111]

Table 7: Categories for the effort level of the questioner

5.1.2 The treatment effect depends on the incentive alignment

According to the model, as shown in equation 6, the effect of a decrease in the exogenous cost of effort is proportional to the degree of incentive alignment between questioner and answerer.

	(1) TWFE	(2) TWFE 2	(3) BJS	(4) BJS 2
Low × after	0.143	-0.0522	0.374**	0.388*
	(0.129)	(0.0693)	(0.140)	(0.153)
MediumLow × after	0.581**	0.401**	0.868***	0.869***
	(0.100)	(0.0543)	(0.164)	(0.188)
MediumHigh × after	0.578**	0.400**	0.884***	0.912***
	(0.103)	(0.0455)	(0.175)	(0.202)
High × after	0.592**	0.413***	0.977***	0.927***
	(0.0709)	(0.0236)	(0.187)	(0.202)
Observations	292919	280407	292846	199564
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 8: 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.

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[23]. Bounties can be considered virtual 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 9.

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

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

[23]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.

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 10. 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 9: 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***	0.652***
	(0.110)	(0.0534)	(0.155)	(0.167)
MediumLow × after	1.235*	1.045*	1.645***	1.088***
	(0.287)	(0.236)	(0.253)	(0.310)
MediumHigh × after	2.296	2.135	2.759***	2.355***
	(0.831)	(0.874)	(0.371)	(0.346)
High × after	3.008***	2.651**	3.477***	2.976***
	(0.268)	(0.209)	(0.439)	(0.460)
Observations	292919	280407	292846	199564
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 10: 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.

5.1.3 Who drives the effect?

The theoretical framework suggests that the size of the effect is proportional to the size of the drop in the cost of language.[24] I do not observe individuals' cost of using English, but, assuming some frictions in switching to the native language website, we would expect that users with a higher cost would be more likely to switch. I then categorize users by the share of posts (i.e. questions or answers) published on a non-English website relative to the total amount of posts published after the native-language website became available. This measure allows to characterize users by the degree they switch to the native-language platform.

To estimate potential heterogeneity on this dimension, I categorize this proxy in four categories, as displayed in table 11. The boundaries of each category are based on the 25^{th}, 50^{th}, and 75^{th} quantiles of the distribution, as displayed in table 5. I then estimate separate treatment effects for each category. 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 12 reports estimates for the four categories. Parameters corresponds to $\{\hat{\beta}_c\}_{\forall c}$ for the TWFE columns, and to $\{\hat{\tau}_c\}_{\forall c}$ for the BJS columns. It is possible to see that the effect is largely driven by users that switch more to the native-language website.

	share of answers not in English in the after-period
Low	[0,0.143]
MediumLow	(0.143,0.426]
MediumHigh	(0.4326,0.875]
High	(0.875,1]

Table 11: Categories for the exogenous cost of using English (boundaries rounded at 2 decimals)

[24] See equation 4.

	(1) TWFE	(2) TWFE 2	(3) BJS	(4) BJS 2
Low × after	0.0988	0.125	0.228	0.212
	(0.114)	(0.102)	(0.182)	(0.193)
MediumLow × after	0.224	0.0889	0.472***	0.217
	(0.122)	(0.106)	(0.141)	(0.146)
MediumHigh × after	0.660*	0.232	0.562***	0.644**
	(0.198)	(0.125)	(0.159)	(0.199)
High × after	1.475***	0.838*	1.883***	2.214***
	(0.142)	(0.174)	(0.271)	(0.304)
Observations	292919	280407	292846	199564
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 12: 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.

5.2 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:

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

Table 13 reports the estimates results. It shows that on average users are significantly more likely to have answers accepted once they can access the website in their native language.

	(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***	0.0203**	0.00873	0.105***	0.105***	0.0931***	0.0705***
	(0.00245)	(0.00240)	(0.00244)	(0.00440)	(0.0168)	(0.0169)	(0.0157)	(0.0165)
Observations	293777	292919	292919	280407	293777	292846	292846	199564
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: 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.

6 Platform's trade-off

Knowledge platforms like Wikipedia and Stack Overflow, that aim to both be global and maximize the quality of the information they are aggregating, have to decide whether they should allow the use of multiple languages. The introduction of multiple versions of the website in different languages has several implications and creates a nontrivial trade-off.

6.1 Benefit: increase in communication effort?

The main analysis discussed in this paper suggests that non-native English-speaking users benefit from a communication cost reduction if allowed to use their native language rather than English. This cost reduction significantly increases users' effort in information transmission. On average, using their native language rather than English, users include 0.66 additional pieces of code. Since the pre-treatment average is 2.71, the effect corresponds to a 24% increase in information quality.

Does this imply that information transmission becomes more effective? In terms of social welfare (and platform's success) what really matters is that users who asked the questions were able to solve their problem. The alternative measure of quality used in the analysis provides some insides in this direction. If Bob asked a question to solve his problem, he can decide to *accept* Alice's answer. Bob is not obliged to do it, so this action can reliably suggest that he solved his problem with Alice's answer. The analysis shows that, when a non-native English speaking user publishes an answer on her native-language website, she is 7% more likely that it gets *accepted*. This is a 20% increase compared to the pre-treatment average (35%). This shows that indeed native-language websites provide a substantial increase in social welfare.

6.2 Benefit: increase in community size?

For the platform, one of the advantages of introducing non-English websites is to potentially reach users who would not be participating otherwise. Table 14 shows that, out of nearly 93K users who published at least one answer/question in one of the non-English websites, 42.8% never registered in the English website. We could guess then that these users would not have joined the platform if their native language would not have become available.

For what concerns users who registered on the English website, if they registered after treatment we cannot disentangle whether they did not participate before because her native language was not available or because of other reasons. Nevertheless, we can observe users that were registered on the English website before being treated but still were not active, meaning that did not contribute any question or answer before treatment. Figure 11 shows the distribution of those users based on what type of participation they made after. It is possible to see that the majority of these users only participated in their native language after treatment.[25] This suggests that the platform may have gained quite some more contributors by introducing websites in other languages. Interestingly, the figure also shows that a relatively large amount of users started participating also in English. This may suggest the presence of positive spillovers.

[25] Russian users do not follow the pattern. This anyway is due to the specificity of the history of the Russian website, as discussed in section 2.2

	After	Before	Not_registered	Tot
SOJ	1579	695	3588	5862
SOP	12178	3386	7800	23364
SOR	23661	279	23352	47292
SOS	7593	3720	5064	16377
Tot	45011	8080	39804	92895

Table 14: Number of active non-native English users who registered in the English website before treatment, after treatment, or did not register. Active means that published at least an answer or question in the non-English websites of the corresponding row.

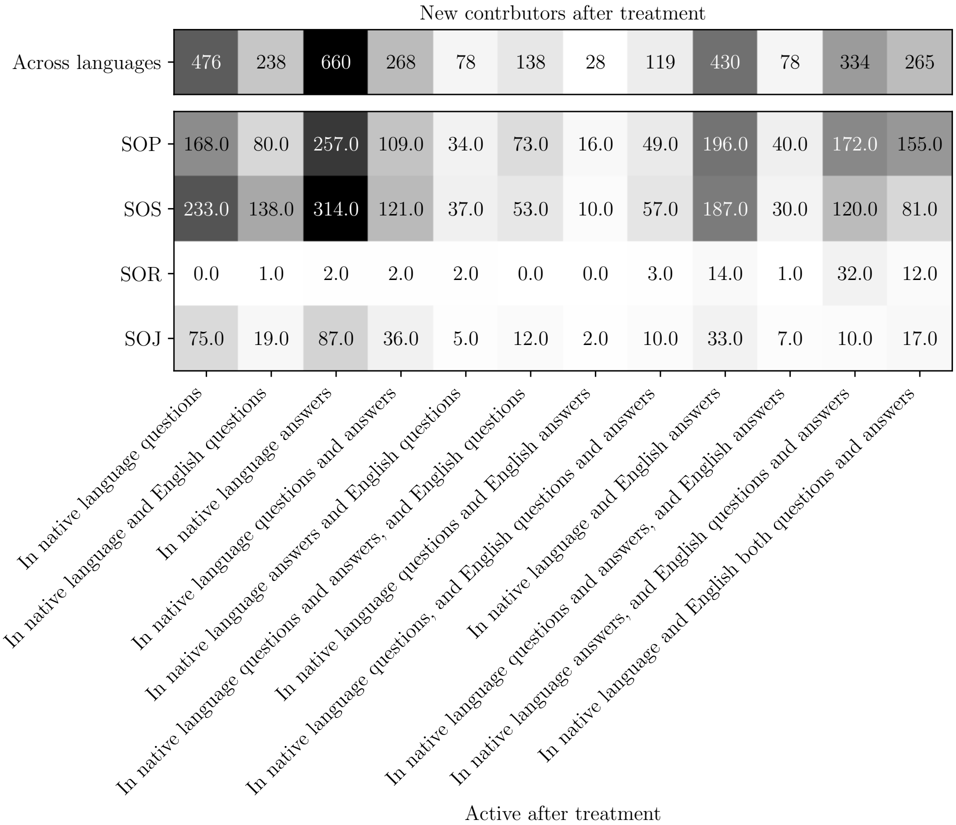

Figure 11: Sample of users who made at least one question and/or answer in a non-English language, who were registered in the English website before treatment, and who did not contribute any question/answer before treatment. Figure reports the number of such users based on what type of contribution they made after treatment.

6.3 Cost: increase in misleading answers?

If communication costs act as a selection device to avoid the participation of users not very expert on the topics of the questions, a reduction in communication costs can lead to more imprecise answers. To show this more precisely through the theoretical framework, let λ''_A be the new level of the cost of language, and λ'_A the initial level, where $\lambda''_A < \lambda'_A$.

Communication effort 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, k_A. 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.

This implies that, if the distribution of expertise across users is the same for different levels of cost of using English, then on average who does not contribute in English before treatment but does contribute in her native language when available, has lower expertise. We should then observe that those users contribute to the native-language website with lower quality answers compared to users who were active in English before treatment.

Figure 12 shows that indeed this is the case. On average, users who were registered but not active before treatment provide lower-quality contributions compared to users who were active before treatment.

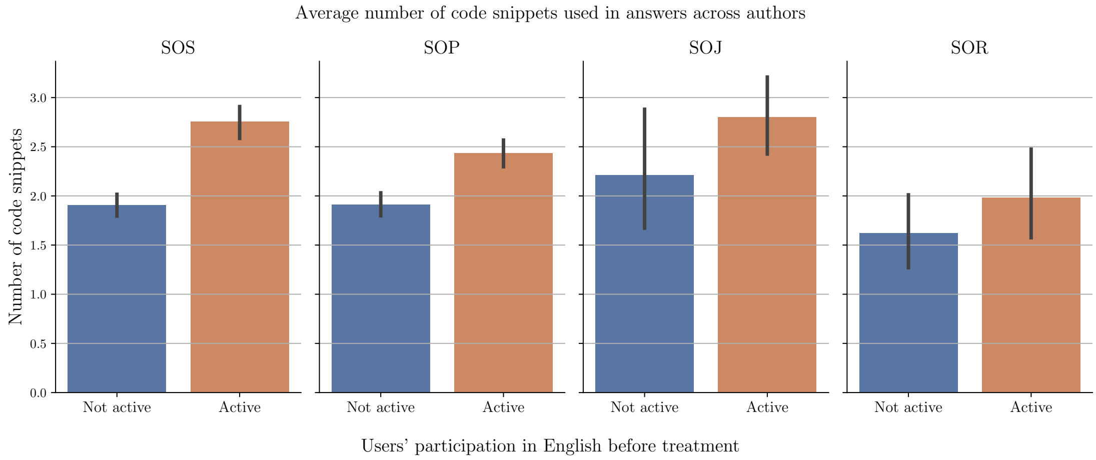

Figure 12: Sample of users who made at least an answer in a non-English language, and who were registered on the English website before treatment. The figure reports the average of the average message quality of each user's contributions, based on whether the user has published answers on the English website before the native-language website became available. Message quality is measured as the number of code snippets appearing in the answer. Vertical black lines are confidence intervals computed via bootstrapping.

6.4 Cost: decrease of communication effort in English?

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.

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 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

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

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 15, in this context the different methodologies provide different results.[26] The preferred approach (i.e. BJS) suggests that there are no significant externalities on the English platform.

	(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*	0.0387	0.203	0.209	0.216	0.203
	(0.150)	(0.129)	(0.129)	(0.0343)	(0.154)	(0.156)	(0.156)	(0.160)
Observations	293777	292919	292919	280407	236495	235574	235574	176512
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 15: 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.

6.5 Cost: reduction in knowledge aggregation?

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.[27]. 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, Blanc and Kubo 2021, Blouin and Dyer 2022). For what concern knowledge platform and Wikipedia in particular, the literature has indeed found that the multiplicity of websites caused the dispersion of information (Bao et al. 2012).

To test if it is the case also for Stack Overflow, I first identify a list of all existing programming languages. I retrieve this list from Wikipedia, which lists 677 programming languages. To avoid confusion with natural languages, let me call the programming languages PLs. I then check if, for each of these PLs, there exists a tag in the Stack

[26] A possible explanation is that, by constraining the sample to only English posts, it is not ensured that users were active both before and after treatment. Some users have in fact fully switched to the native-language platform after treatment.

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

Overflow websites. In Stack Overflow, tags are used to categorize the content of the questions. This implies that, if a tag exists, then at least one question has addressed that topic. If a tag exists in some languages, but not in others, it means that only the community of that language has addressed that topic.

Table 16 shows the number of PLs that appeared in 0, 1, 2, 3, or 4 languages, where the languages are Spanish, Portuguese, Russian, and Japanese, and whether they also appeared in English. It shows that out of the 677 PLs, only 28 of them appear in all 5 languages (including English). Out of the 247 PLs discussed in at least one language, 219 are discussed only in some of the languages, meaning that the information is not accessible for users not speaking those languages. In addition, 83 PLs are discussed in at least two languages, suggesting that there could be efficiency gains if everyone would speak the same language. Finally, 12 PLs are discussed only in languages different from English. This suggests the potential risk that the implementation of additional languages has reduced the variety of information in English.

Number of non-English languages with the tag Whether tag is in English site	0.0	1.0	2.0	3.0	4.0
0.0	430	8	3	1	
1.0	152	29	17	9	28

Table 16: Number of programming languages for which at least a question has been made in 1, 2, 3, or 4 languages. The 4 languages are Spanish, Portuguese, Russian, and Japanese. Rows split the sample based on whether the tag appears in the English website (1) or not (0)

7 Conclusion

This paper studies the trade-off faced by knowledge platforms when deciding to make their website available in either one or multiple languages.

It shows that the benefits of allowing contributions in multiple languages are substantial, mainly because it reduces communication costs for the users native to those languages. On one side, I show that at least 42% of those users were unlikely to participate if their native-language website was not available. On the other, users increase by 24% their communication effort after their native language becomes available. This effect is driven by users that, after their native language became available, have switched the most to it, reducing contributions in English. These users are likely to be the ones with the highest cost of using English, and who then faced the largest drop in communication costs by using their native language. The increase in effort due to a reduction in the cost of language is positively correlated to the questioners' effort and incentives. When answering in their native language, users increase effort by up to 34% if the questions are in the top quartile by quality, and up to 110% if they are highly incentivized via virtual remuneration.

The paper then shows that there are no clear negative externalities for the English website: the users who keep contributing to the English website after their native-language website becomes available do not significantly reduce the effort in English. For what concerns the extensive margin instead, a substantial amount of users stop participating in English when their native language is available. These anyway are users who were not very active anyway: users that were contributing a lot in English kept a high level of participation.

According to this evidence, it seems advisable to introduce websites in multiple languages, as it increases the community size and the quality of the information collected. Nevertheless, the paper shows also some drawbacks. First of all, the new inflow of participation induced by the availability of additional languages is characterized by lower-quality contributions, which is reducing the overall improvement in information quality. This is justifiable by the fact that a high cost of language acts as a barrier to participation for inexpert users. Second, there is naturally a decrease in efficiency in information aggregation. If the same topic is addressed in more than one language it means that multiple users spent time and effort to potentially provide the same piece of information. At the same time, if some information is provided only in some languages but not others, then some users are not able to access it. Both issues would be solved by imposing a single language. It follows that, if the communities of non-English speakers are small and few people would benefit from multiple languages, a single language is preferable. This anyway would raise ethical concerns, as it would exclude minorities or constrain their access to the platform.

Overall is not clear what is the optimal strategy, which then depends on the long-term objective of the platform (e.g. how global it wants to be) and the size of the communities using given languages. It seems wise for Stack Overflow to have implemented additional websites for only some of the most common languages outside English.

While the analysis is specific to the context of Stack Overflow, the results may contribute to different environments. A large literature has addressed communication costs as a major constraint to efficient economic activities, but, to my knowledge, it has not quantified the problem. This is relevant anyway for a variety of decision-makers. To give a few examples, when firms need to form teams of employees of different nationalities, they need to assess the advantages of pairing co-workers of the same nationality (Lyons 2017). In defining the hierarchy structure of the company, managers need to evaluate the advantage of hiring *translators* or imposing the same language across teams (Crémer et al. 2007). Finally, national states may want to understand the exact benefit of imposing a homogeneous language before taking initiative toward centralization.

This paper is silent on how organizations could compensate and alleviate part of the trade-off using external technologies. For instance, the development of live translations and search engines that allow for searches across languages may solve the trade-off. Many issues anyway may reduce the benefits of those technologies. For example, a lot of expressions and concepts require a complete rewriting to convey the same message in different languages, something that only human translators can achieve. Future research should then be devoted to understanding to what extent existing or potential future technologies could be instrumental. In addition, future work should be devoted to investigating the external validity of these findings in the context of face-to-face communications with personal interactions.

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, 34

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. 5

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. 18

Blanc, G. and M. Kubo (2021). Schools, language, and nations: Evidence from a natural experiment in France. *Working paper*. 5, 34

Blouin, A. and J. Dyer (2022). How cultures converge: An empirical investigation of trade and linguistic exchange. *Working paper*. 34

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. 5

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*. 4, 22, 23

Callaway, B. and P. H. C. Sant'Anna (2020, December). Difference-in-Differences with multiple time periods. *Journal of Econometrics*. 4, 22

Calvó-Armengol, A., J. de Martí, and A. Prat (2015). Communication and influence. *Theoretical Economics 10*, 649–690. 7

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, 5, 36

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. 4, 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, 5, 34

Ginsburgh, V. and S. Weber (2020, June). The Economics of Language. *Journal of Economic Literature 58*(2), 348–404. 18

Goodman-Bacon, A. (2021, December). Difference-in-differences with variation in treatment timing. *Journal of Econometrics 225*(2), 254–277. 22

Guillouët, L., A. K. Khandelwal, R. Macchiavello, and M. Teachout (2021). Language barriers in multinationals and knowledge transfers. *Working paper*. 5

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. 5

Lohmann, J. (2011, February). Do language barriers affect trade? *Economics Letters 110*(2), 159–162. 5

Lyons, E. (2017, July). Team Production in International Labor Markets: Experimental Evidence from the Field. *American Economic Journal: Applied Economics 9*(3), 70–104. 18, 36

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. 5

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*. 4, 22

Zhang, X. M. and F. Zhu (2011, June). Group size and incentives to contribute: A natural experiment at chinese wikipedia. *The American Economic Review 101*(4), 1601–1615. 4

Appendix A Additional details on Stack Overflow

A.1 The introduction of new websites

The creation of new websites of Stack Overflow follows a specific process. The main objective is to ensure, before the launch, a sufficiently active community base that will guarantee the growth and the sustainability of the website in the long run. First of all the website is proposed in an ad-hoc platform called *Area 51*, where users registered can support the proposal and start publishing questions and answers. If the website idea receives enough attention and contributions, then it proceeds to the *beta* period, it gets its URL, and it is accessible as an independent site. The *beta* period is split into two steps: first, in the so-called *private beta*, only users that were active in supporting it in the early stage can contribute. Then, when it becomes *public beta*, everyone can register and contribute. Once all features are implemented, the website is said to *graduate*, entering its final stage. At each stage, the incentive system may slightly vary. For example, some *privileges* are reachable with different with different amounts of points, generally being lower requirements in earlier stages.

Data is available starting from the *private beta* period. Table 17 reports the dates for the start of each stage for websites in different languages.

Appendix B Details about the theoretical framework

platform	proposal	private beta	public beta	graduation
SO		01/08/2008	-	15/09/2008
SO - R	01/06/2012	27/03/2015	27/03/2015	11/12/2015
SO - J	-	29/09/2014	16/12/2014	[not graduated]
SO - S	02/08/2012	01/12/2015	15/12/2015	17/5/2017
SO - P	05/11/2010	12/12/2013	29/01/2014	15/5/2015

Table 17: Dates in which the platforms passed the different development stages. SO correspond to Stack Overflow in English, while the initials R, J, S, P stand for Russian, Japanese, Spanish, and Portuguese, respectively.

B.1 Second stage

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

where the last equality holds because of Bayes Normal updating.

B.2 First stage

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

Given the action expected to be chosen by Bob, the problem rewrites as:

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

and

$$\frac{\gamma E_Q^2}{(E_Q E_A + E_Q s + E_A s)^2} = C_A^2$$

$$(E_Q E_A + E_Q s + E_A s)^2 = \frac{\gamma E_Q^2 k_A^2}{\lambda_A^2}$$

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

The best response is then given by:

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

Appendix C Additional details about the data and the measures

C.1 Quality measure

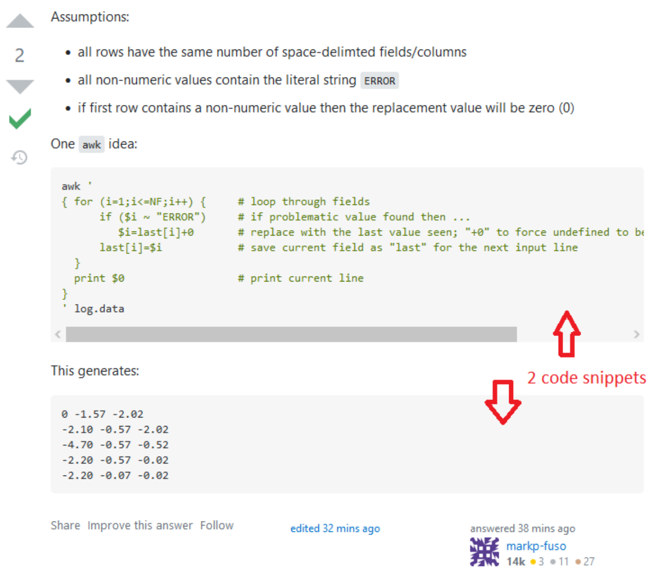

Figure 13: Example of an answer in Stack Overflow where the number of snippets (i.e. the proxy for quality) is equal to 2.

C.2 Participation choices

Appendix D Robustness for DiD analysis

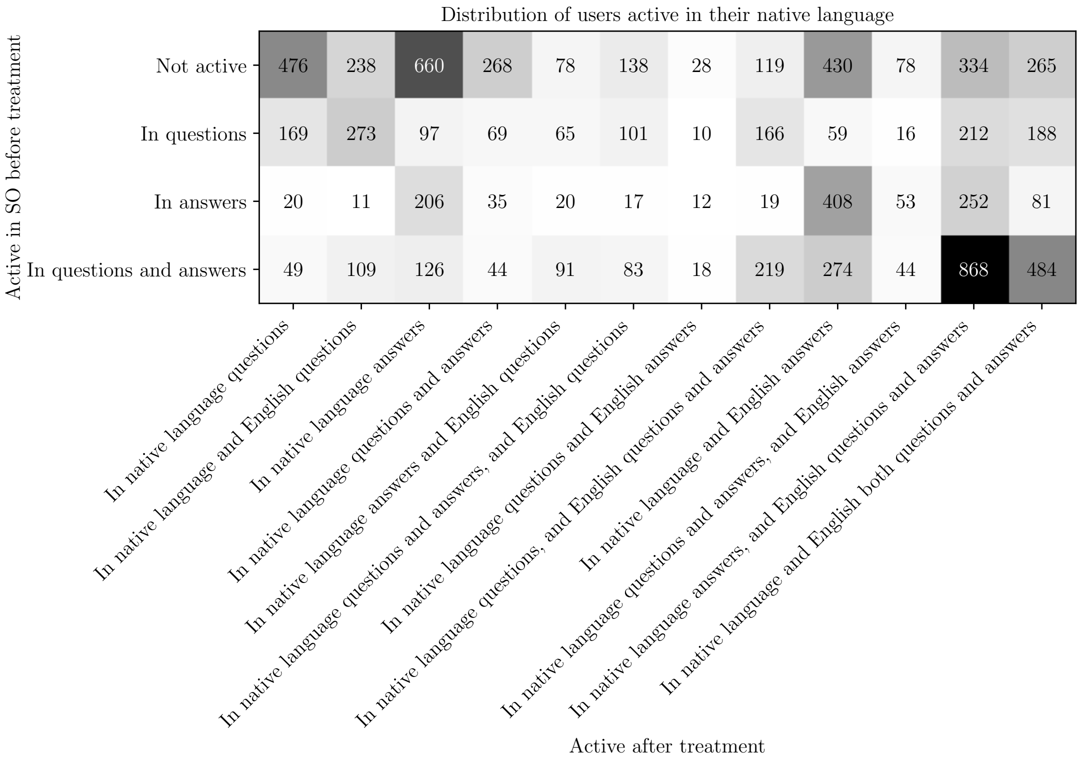

Figure 14: Distribution of users based on what type of contributions they have made. The sample conditions for 1) The user must have registered in Stack Overflow (English) before being treated, and 2) The user participated with at least one question/ answer in a non-English website. The y-axis identifies contributions made before being treated (i.e. in English only), while the x-axis the contributions made after treatment.

D.1 removing answers with 0 pieces of code

Table 18 reports regression results comparable to the estimation in table 6 after dropping all answers with zero pieces of code and selecting users that, given the remaining answers, where active both before and after treatment.

	(1) TWFE	(2) TWFE 1	(3) TWFE 2	(4) TWFE 3	(5) BJS	(6) BJS 1	(7) BJS 2	(8) BJS 3
after	0.442^{*}	0.434^{*}	0.434^{*}	0.209	0.838^{***}	0.864^{***}	0.866^{***}	0.910^{***}
	(0.141)	(0.147)	(0.147)	(0.0743)	(0.185)	(0.185)	(0.186)	(0.170)
Observations	228244	227818	227818	218908	228244	227812	227812	152843
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 18: Baseline Regressions' estimates where the dependent variable is the number of pieces of code, after dropping all answers with zero 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.