Appendix for: Bregolin, Jacopo, *Communication Quality and the Cost of Language: Evidence from Stack Overflow*

Appendix. Additional material

A. Details about Stack Overflow

A.1. The Introduction of New Websites

The creation of new Stack Overflow websites follows a specific process. The main objective is to ensure, before the launch, a sufficiently active community base that will guarantee the website's growth and long-term sustainability. First of all, the website is proposed in an ad-hoc platform called *Area 51* where registered users can support the proposal and start publishing questions and answers. If the website idea receives enough attention and contributions, it proceeds to the *beta* period, gets its URL, and becomes accessible as an independent site. The *beta* period is split into two steps. First, in the so-called *private beta*, only users who actively supported 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 vary slightly. For example, some *privileges* are reachable with different amounts of points, with generally lower requirements in earlier stages.[44]

Data is available starting from the *private beta* period. Table EC.1 reports the dates for the start of each stage for websites in different languages.

Platform	Proposal	Private beta	Public beta	Graduation
English		01/08/2008	-	15/09/2008
Russian	01/06/2012	27/03/2015	27/03/2015	11/12/2015
Japanese	-	29/09/2014	16/12/2014	[not graduated]
Spanish	02/08/2012	01/12/2015	15/12/2015	17/5/2017
Portuguese	05/11/2010	12/12/2013	29/01/2014	15/5/2015

Table EC.1 Dates in which the platforms passed the different development stages.

B. Details About the Theoretical Framework

I solve the model by backward induction.

B.1. Solving the Model: Second Stage

In the second stage, the questioner observes the message m and chooses the action to address his problem and maximise his expected utility based on the message received. Note that the expectation is taken with respect to θ.

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

[44] https://meta.stackexchange.com/questions/58587/reputation-requirements-compared

$$\begin{aligned}
&\iff a^* \equiv \arg\max_a -a^2 - \mathbb{E}[\theta^2|m] + 2a\mathbb{E}[\theta|m] + C_Q^2\Phi_Q \\
&\iff -2a^* + 2\mathbb{E}[\theta|m] = 0 \\
&\iff a^* = \mathbb{E}[\theta|m].
\end{aligned}$$

Since both θ and m are normally distributed, we can use Bayesian updating for normal random variables to compute the expectation.[45][46]

Recall that $m = \theta + \varepsilon + \eta$, $\theta \sim \mathcal{N}(0, \frac{1}{s})$, $\varepsilon \sim \mathcal{N}(0, \frac{1}{E_Q})$, $\eta \sim \mathcal{N}(0, \frac{1}{E_A})$, and $\{\theta, \varepsilon, \eta\}$ are independent. It follows that $(\theta, m) \sim \mathcal{N}(\mu, \Sigma)$ with:

$$\mu = \begin{bmatrix} \mu_\theta \\ \mu_m \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \text{and} \quad \Sigma = \begin{bmatrix} \Sigma_{\theta,\theta} & \Sigma_{\theta,m} \\ \Sigma_{m,\theta} & \Sigma_{m,m} \end{bmatrix} = \Sigma = \begin{bmatrix} \frac{1}{s} & \frac{1}{s} \\ \frac{1}{s} & \frac{1}{s} + \frac{1}{\Phi_Q} + \frac{1}{\Phi_A} \end{bmatrix}.$$

We then have that:

$$\begin{aligned}
a^* &= \mathbb{E}[\theta|m] \\
&= \mu_\theta + \frac{\Sigma_{\theta,m}}{\Sigma_{m,m}}(m - \mu_m) \\
&= \frac{\frac{1}{s}}{\frac{1}{s} + \frac{1}{E_Q} + \frac{1}{E_Q}} m \\
&= \beta m \quad \text{with} \quad \beta \equiv \frac{\Phi_Q\Phi_A}{\Phi_Q\Phi_A + \Phi_Q s + \Phi_A s}.
\end{aligned}$$

B.2. Solving the Model: First Stage

The answerer chooses the quality level to maximise her expected utility, where the expectation is with respect to θ, ε, and η, as the message is not yet realised.

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

Based on the action expected to be chosen by Bob, the problem rewrites as follows:

$$\begin{aligned}
&\max_{\Phi_A \geq 0} -\gamma\mathbb{E}[(\beta m - \theta)^2] - C_A^2\Phi_A \\
\iff &\max_{\Phi_A \geq 0} -\gamma\mathbb{E}[\beta m - \theta]^2 - \gamma\mathbb{V}[\beta m - \theta] - C_A^2\Phi_A \quad \text{(by property of the variance)} \\
\iff &\max_{\Phi_A \geq 0} -\gamma\mathbb{V}[\beta m - \theta] - C_A^2\Phi_A \quad \text{(since } m \text{ and } \theta \text{ have zero mean)} \\
\iff &\max_{\Phi_A \geq 0} -\gamma\left(\beta^2\mathbb{V}[m] + \mathbb{V}[\theta] - 2\beta\mathbb{V}[\theta]\right) - C_A^2\Phi_A \\
&\qquad \text{Note that } \mathbb{V}[m] = \frac{\Phi_Q\Phi_A + \Phi_Q s + \Phi_A s}{s\Phi_Q\Phi_A} = \frac{1}{\beta s} \text{ since } \beta \equiv \frac{\Phi_Q\Phi_A}{\Phi_Q\Phi_A + \Phi_Q s + \Phi_A s}, \text{ so:} \\
\iff &\max_{\Phi_A \geq 0} -\gamma\left(\beta^2\frac{1}{\beta s} + \frac{1}{s} - 2\beta\frac{1}{s}\right) - C_A^2\Phi_A \\
\iff &\max_{\Phi_A \geq 0} -\gamma\left(\beta\frac{1}{s} + \frac{1}{s} - 2\beta\frac{1}{s}\right) - C_A^2\Phi_A \\
\iff &\max_{\Phi_A \geq 0} -\gamma\left(\frac{1}{s}(1-\beta)\right) - C_A^2\Phi_A = -\gamma\frac{1}{s} + \gamma\frac{1}{s}\beta - C_A^2\Phi_A.
\end{aligned}$$

[45] Note that to avoid introducing cumbersome notation, I am using m to identify both the random variable and its realisation.

[46] The result is reported in Vives (2008)'s technical appendix (section 10.2.1, page 376) and shown in DeGroot (1970). Consider two normal random variables $(\theta, s) \sim \mathcal{N}(\mu, \Sigma)$ such that the mean vector and the variance-covariance matrix correspond to $\mu = \begin{bmatrix} \mu_\theta \\ \mu_s \end{bmatrix}$ and $\Sigma = \begin{bmatrix} \Sigma_{\theta,\theta} & \Sigma_{\theta,s} \\ \Sigma_{s,\theta} & \Sigma_{s,s} \end{bmatrix}$. Then, the conditional density of θ given s is $(\theta \mid s) \sim \mathcal{N}\left(\mu_\theta + \Sigma_{\theta,s}\Sigma_{s,s}^{-1}(s - \mu_s), \Sigma_{\theta,\theta} - \Sigma_{\theta,s}\Sigma_{s,s}^{-1}\Sigma_{s,\theta}\right)$.

By first-order condition, the best response quality level satisfies:

$$\begin{aligned}
\frac{\partial}{\partial \Phi_A}\left(-\gamma\frac{1}{s}+\gamma\frac{1}{s}\frac{\Phi_Q\Phi_A}{\Phi_Q\Phi_A+\Phi_Q s+\Phi_A s}-C_A^2\Phi_A\right)&=0\\
\frac{\gamma\Phi_Q^2}{(\Phi_Q\Phi_A+\Phi_Q s+\Phi_A s)^2}&=C_A^2\equiv\left(\frac{\lambda_A}{k_A}\right)^2\\
(\Phi_Q\Phi_A+\Phi_Q s+\Phi_A s)^2&=\frac{\gamma\Phi_Q^2k_A^2}{\lambda_A^2}\\
\Phi_Q\Phi_A+\Phi_Q s+\Phi_A s&=\sqrt{\frac{\gamma\Phi_Q^2k_A^2}{\lambda_A^2}}\\
\Phi_A(\Phi_Q+s)&=\frac{\sqrt{\gamma}\Phi_Q k_A}{\lambda_A}-\Phi_Q s\\
\Phi_A(\Phi_Q+s)&=\frac{\Phi_Q(\sqrt{\gamma}k_A-s\lambda_A)}{\lambda_A}.
\end{aligned}$$

The best response is then given by:

$$BR_A(\Phi_Q)=\frac{\Phi_Q(\sqrt{\gamma}k_A-s\lambda_A)}{\lambda_A(\Phi_Q+s)}. \tag{EC.1}$$

Note that in the first-order condition, a second solution of the quadratic equation is excluded as it takes only negative values, which is not allowed by the model assumptions (i.e. the quality level is weakly positive).

B.3. Comparative Statics

Under the condition that quality is positive (i.e. under the assumption that $\sqrt{\gamma}k_A > s\lambda_A)$), and everything else held constant, a marginal increase in language cost affects the quality choice in the following way:

$$\begin{aligned}
\frac{\partial BR_A(\Phi_Q)}{\partial\lambda_A}&=\frac{\partial}{\partial\lambda_A}\left(\frac{\Phi_Q(\sqrt{\gamma}k_A-s\lambda_A)}{\lambda_A(\Phi_Q+s)}\right)=\frac{\partial}{\partial\lambda_A}\left(\frac{\Phi_Q\sqrt{\gamma}k_A}{\lambda_A(\Phi_Q+s)}-\frac{\Phi_Q s}{\Phi_Q+s}\right)\\
&=-\frac{(\Phi_Q+s)\Phi_Q\sqrt{\gamma}k_A}{\lambda_A^2(\Phi_Q+s)^2}\\
&=-\frac{\Phi_Q\sqrt{\gamma}k_A}{\lambda_A^2(\Phi_Q+s)}<0.
\end{aligned}$$

The second degree effect is given by:

$$\begin{aligned}
\frac{\partial BR_A(\Phi_Q)}{\partial\lambda_A^2}&=\frac{\partial}{\partial\lambda_A}\left(-\frac{\Phi_Q\sqrt{\gamma}k_A}{\lambda_A^2(\Phi_Q+s)}\right)\\
&=-\frac{2\lambda_A(\Phi_Q+s)(-\Phi_Q\sqrt{\gamma}k_A)}{\lambda_A^4(\Phi_Q+s)^2}\\
&=\frac{2\Phi_Q\sqrt{\gamma}k_A}{\lambda_A^3(\Phi_Q+s)}>0.
\end{aligned}$$

The change in quality due to a marginal change in the language cost depends on the quality of the question:

$$\begin{aligned}
\frac{\partial BR_A(\Phi_Q)}{\partial\lambda_A\partial\Phi_Q}&=\frac{\partial}{\partial\Phi_Q}\left(-\frac{\Phi_Q\sqrt{\gamma}k_A}{\lambda_A^2(\Phi_Q+s)}\right)\\
&=-\frac{\sqrt{\gamma}k_A(\lambda_A^2(\Phi_Q+s))-\lambda_A^2(\Phi_Q\sqrt{\gamma}k_A)}{\lambda_A^4(\Phi_Q+s)^2}\\
&=-\frac{\sqrt{\gamma}k_A s}{\lambda_A^2(\Phi_Q+s)^2}<0.
\end{aligned}$$

The change in quality due to a marginal change in the language cost depends on the incentive degree:

$$\begin{aligned}\frac{\partial BR_A(\Phi_Q)}{\partial\lambda_A\partial\gamma} &= \frac{\partial}{\partial\gamma}\left(-\frac{\Phi_Q\sqrt{\gamma}k_A}{\lambda_A^2(\Phi_Q+s)}\right)\\ &= \left(-\frac{\Phi_Q k_A}{\lambda_A^2(\Phi_Q+s)}\right)\frac{1}{2\sqrt{\gamma}}\\ &= -\frac{\Phi_Q k_A}{2\lambda_A^2(\Phi_Q+s)\sqrt{\gamma}} < 0.\end{aligned}$$

The change in quality due to a change in γ is given by:

$$\begin{aligned}\frac{\partial BR_A(\Phi_Q)}{\partial\gamma} &= \frac{\partial}{\partial\gamma}\left(\frac{\Phi_Q(\sqrt{\gamma}k_A - s\lambda_A)}{\lambda_A(\Phi_Q+s)}\right) = \frac{\partial}{\partial\gamma}\left(\frac{\Phi_Q\sqrt{\gamma}k_A}{\lambda_A(\Phi_Q+s)} - \frac{\Phi_Q s}{\Phi_Q+s}\right)\\ &= \frac{\Phi_Q k_A}{\lambda_A(\Phi_Q+s)}\frac{1}{2\sqrt{\gamma}} > 0.\end{aligned}$$

The change in quality due to a change in k_A is given by:

$$\begin{aligned}\frac{\partial BR_A(\Phi_Q)}{\partial k_A} &= \frac{\partial}{\partial k_A}\left(\frac{\Phi_Q(\sqrt{\gamma}k_A - s\lambda_A)}{\lambda_A(\Phi_Q+s)}\right) = \frac{\partial}{\partial k_A}\left(\frac{\Phi_Q\sqrt{\gamma}k_A}{\lambda_A(\Phi_Q+s)} - \frac{\Phi_Q s}{\Phi_Q+s}\right)\\ &= \frac{\Phi_Q\sqrt{\gamma}}{\lambda_A(\Phi_Q+s)} > 0.\end{aligned}$$

B.4. Details and Theoretical Derivation for how *Old Joiners* have Higher Expertise than *New Joiners*

Section 4.3.2 provides theoretical predictions that rely on the result that *old joiners* have higher expertise than *new joiners*. This section aims to provide additional details and show that this relationship holds even if answerers differ in their cost of using English.

Let $j \in (0,1)$ index answerers participating on a topic ω. Let k_j be the expertise of user j on the topic ω.[47] In addition, let $\lambda \in \Lambda$ define the cost of using English and λ' be the cost of using the native language (assumed fixed across users), with $\lambda' < \lambda\ \forall\lambda \in \Lambda$. Moreover, let NJ identify the set of *new joiners*, and OJ the set of *old joiners*, such that:

$$\begin{aligned} j \in NJ \quad &\text{if} \quad \frac{s\lambda'}{\sqrt{\gamma}} < k_j < \frac{s\lambda_j}{\sqrt{\gamma}}\\ j \in OJ \quad &\text{if} \quad k_j > \frac{s\lambda_j}{\sqrt{\gamma}},\end{aligned}$$

where k_j is j's expertise on the topic ω and γ and s are assumed to be constants. Let $f(k)$ and $F(k)$ define, respectively, the probability density function and the cumulative density function over levels of expertise and across answerers. Assume that these distributions are fixed across different cost levels for using English (λ).

B.4.1. Case where all Answerers have the Same Cost of Using English. This is the case discussed in the main text. Since answerers are only allowed to differ in their expertise (k_j), we have that

$$\mathbb{E}\left[k_j|\omega, j \in OJ\right] \geq \mathbb{E}\left[k_j|\omega, j \in NJ\right]$$

by direct implication of the definition of *old joiners* and *new joiners*.

[47] For the sake of clarity, compared to the main text, I have dropped the subscript A on all notation that refers to expertise and language cost, as it is not needed for the exposition of this proof.

B.4.2. Case with Two Possible Cost Levels for the use of English. To provide a more tractable example, consider the case where there are two types of answerers: one has a high cost of using English ($\overline{\lambda}$) while the other has a low cost of using English ($\underline{\lambda}$). Figure EC.1 represents the setting in the particular case that f is the density of a normal distribution. On the English site, contributions would come from a share $1-F\left(\frac{s\underline{\lambda}}{\sqrt{\gamma}}\right)$ of low-cost answerers and $1-F\left(\frac{s\overline{\lambda}}{\sqrt{\gamma}}\right)$ of high-cost answerers. On the native language site, additional contributions would come from a share $F\left(\frac{s\underline{\lambda}}{\sqrt{\gamma}}\right)-F\left(\frac{s\lambda'}{\sqrt{\gamma}}\right)$ of low-cost answerers and a share $F\left(\frac{s\overline{\lambda}}{\sqrt{\gamma}}\right)-F\left(\frac{s\lambda'}{\sqrt{\gamma}}\right)$ of high-cost answerers. For the sake of readability, I use κ, $\overline{\kappa}$, $\underline{\kappa}$, and κ' to mean $\frac{s\underline{\lambda}}{\sqrt{\gamma}}$, $\frac{s\overline{\lambda}}{\sqrt{\gamma}}$, $\frac{s\underline{\lambda}}{\sqrt{\gamma}}$, and $\frac{s\lambda'}{\sqrt{\gamma}}$, respectively. We then have that:

$$
\begin{aligned}
\mathbb{E}\left[k_j|\omega, j\in OJ\right] &= \frac{\int_{\underline{\kappa}}^{\infty} kf(k)dk+\int_{\overline{\kappa}}^{\infty} kf(k)dk}{[1-F(\underline{\kappa})]+[1-F(\overline{\kappa})]}\\
&= \frac{1-F(\underline{\kappa})}{[1-F(\underline{\kappa})]+[1-F(\overline{\kappa})]}\frac{\int_{\underline{\kappa}}^{\infty} kf(k)dk}{1-F(\underline{\kappa})}+\frac{1-F(\overline{\kappa})}{[1-F(\underline{\kappa})]+[1-F(\overline{\kappa})]}\frac{\int_{\overline{\kappa}}^{\infty} kf(k)dk}{1-F(\overline{\kappa})}\\
&= \frac{1-F(\underline{\kappa})}{[1-F(\underline{\kappa})]+[1-F(\overline{\kappa})]}\left(\frac{\int_{\underline{\kappa}}^{\overline{\kappa}} kf(k)dk}{1-F(\underline{\kappa})}+\frac{\int_{\overline{\kappa}}^{\infty} kf(k)dk}{1-F(\underline{\kappa})}\right)+\frac{1-F(\overline{\kappa})}{[1-F(\underline{\kappa})]+[1-F(\overline{\kappa})]}\frac{\int_{\overline{\kappa}}^{\infty} kf(k)dk}{1-F(\overline{\kappa})}\\
&= \frac{1-F(\underline{\kappa})}{[1-F(\underline{\kappa})]+[1-F(\overline{\kappa})]}\left(\frac{F(\overline{\kappa})-F(\underline{\kappa})}{1-F(\underline{\kappa})}\frac{\int_{\underline{\kappa}}^{\overline{\kappa}} kf(k)dk}{F(\overline{\kappa})-F(\underline{\kappa})}+\frac{1-F(\overline{\kappa})}{1-F(\underline{\kappa})}\frac{\int_{\overline{\kappa}}^{\infty} kf(k)dk}{1-F(\overline{\kappa})}\right)\\
&+ \frac{1-F(\overline{\kappa})}{[1-F(\underline{\kappa})]+[1-F(\overline{\kappa})]}\frac{\int_{\overline{\kappa}}^{\infty} kf(k)dk}{1-F(\overline{\kappa})}\\
&= \frac{F(\overline{\kappa})-F(\underline{\kappa})}{[1-F(\underline{\kappa})]+[1-F(\overline{\kappa})]}\frac{\int_{\underline{\kappa}}^{\overline{\kappa}} kf(k)dk}{F(\overline{\kappa})-F(\underline{\kappa})}+\frac{2[1-F(\overline{\kappa})]}{[1-F(\underline{\kappa})]+[1-F(\overline{\kappa})]}\frac{\int_{\overline{\kappa}}^{\infty} kf(k)dk}{1-F(\overline{\kappa})}\\
&= \frac{[F(\overline{\kappa})-F(\underline{\kappa})]B+2[1-F(\overline{\kappa})]C}{[F(\overline{\kappa})-F(\underline{\kappa})]+2[1-F(\overline{\kappa})]},
\end{aligned}
$$

where B is the average expertise for levels between $\underline{\kappa}$ and $\overline{\kappa}$ while C is the average expertise for levels above $\overline{\kappa}$ (i.e. of areas $\mathcal{B}$ and $\mathcal{C}$ in figure EC.1). For the *new joiners*, we have that:

$$
\begin{aligned}
\mathbb{E}\left[k_j|\omega, j\in NJ\right] &= \frac{\int_{\kappa'}^{\underline{\kappa}} kf(k)dk+\int_{\kappa'}^{\overline{\kappa}} kf(k)dk}{[F(\underline{\kappa})-F(\kappa')]+[F(\overline{\kappa})-F(\kappa')]}\\
&= \frac{F(\underline{\kappa})-F(\kappa')}{[F(\underline{\kappa})-F(\kappa')]+[F(\overline{\kappa})-F(\kappa')]}\frac{\int_{\kappa'}^{\underline{\kappa}} kf(k)dk}{F(\underline{\kappa})-F(\kappa')}+\frac{F(\overline{\kappa})-F(\kappa')}{[F(\underline{\kappa})-F(\kappa')]+[F(\overline{\kappa})-F(\kappa')]}\frac{\int_{\kappa'}^{\overline{\kappa}} kf(k)dk}{F(\overline{\kappa})-F(\kappa')}\\
&= \frac{F(\underline{\kappa})-F(\kappa')}{[F(\underline{\kappa})-F(\kappa')]+[F(\overline{\kappa})-F(\kappa')]}\frac{\int_{\kappa'}^{\underline{\kappa}} kf(k)dk}{F(\underline{\kappa})-F(\kappa')}\\
&+ \frac{F(\overline{\kappa})-F(\kappa')}{[F(\underline{\kappa})-F(\kappa')]+[F(\overline{\kappa})-F(\kappa')]}\left(\frac{\int_{\kappa'}^{\underline{\kappa}} kf(k)dk}{F(\overline{\kappa})-F(\kappa')}+\frac{\int_{\underline{\kappa}}^{\overline{\kappa}} kf(k)dk}{F(\overline{\kappa})-F(\kappa')}\right)\\
&= \frac{F(\underline{\kappa})-F(\kappa')}{[F(\underline{\kappa})-F(\kappa')]+[F(\overline{\kappa})-F(\kappa')]}\frac{\int_{\kappa'}^{\underline{\kappa}} kf(k)dk}{F(\underline{\kappa})-F(\kappa')}\\
&+ \frac{F(\overline{\kappa})-F(\kappa')}{[F(\underline{\kappa})-F(\kappa')]+[F(\overline{\kappa})-F(\kappa')]}\left(\frac{F(\underline{\kappa})-F(\kappa')}{F(\overline{\kappa})-F(\kappa')}\frac{\int_{\kappa'}^{\underline{\kappa}} kf(k)dk}{F(\underline{\kappa})-F(\kappa')}+\frac{F(\overline{\kappa})-F(\underline{\kappa})}{F(\overline{\kappa})-F(\kappa')}\frac{\int_{\underline{\kappa}}^{\overline{\kappa}} kf(k)dk}{F(\overline{\kappa})-F(\underline{\kappa})}\right)\\
&= \frac{2[F(\underline{\kappa})-F(\kappa')]}{[F(\underline{\kappa})-F(\kappa')]+[F(\overline{\kappa})-F(\kappa')]}\frac{\int_{\kappa'}^{\underline{\kappa}} kf(k)dk}{F(\underline{\kappa})-F(\kappa')}+\frac{F(\overline{\kappa})-F(\underline{\kappa})}{[F(\underline{\kappa})-F(\kappa')]+[F(\overline{\kappa})-F(\kappa')]}\frac{\int_{\underline{\kappa}}^{\overline{\kappa}} kf(k)dk}{F(\overline{\kappa})-F(\underline{\kappa})}\\
&= \frac{2[F(\underline{\kappa})-F(\kappa')]A+[F(\overline{\kappa})-F(\underline{\kappa})]B}{2[F(\underline{\kappa})-F(\kappa')]+[F(\overline{\kappa})-F(\underline{\kappa})]},
\end{aligned}
$$

where A is the average expertise for levels below $\underline{\kappa}$ (i.e. of area $\mathcal{A}$ in figure EC.1).

Note that $B < C$. Indeed:

$$\begin{aligned}\int_{\underline{\kappa}}^{\infty} kf(k)dk &= \overline{\kappa} + \int_{\underline{\kappa}}^{\overline{\kappa}} (k-\overline{\kappa})f(k)dk + \int_{\overline{\kappa}}^{\infty} (k-\overline{\kappa})f(k)dk \\ &\leq \overline{\kappa} + \int_{\overline{\kappa}}^{\infty} (k-\overline{\kappa})f(k)dk \\ &= \overline{\kappa} + [1-F(\overline{\kappa})]\frac{\int_{\overline{\kappa}}^{\infty}(k-\overline{\kappa})f(k)dk}{1-F(\overline{\kappa})} \\ &\leq \overline{\kappa} + \frac{\int_{\overline{\kappa}}^{\infty}(k-\overline{\kappa})f(k)dk}{1-F(\overline{\kappa})} = C,\end{aligned}$$

where the first inequality holds as it removes something weakly negative, while the second inequality holds because $0 \leq [1-F(\overline{\kappa})] \leq 1$ and $\int_{\overline{\kappa}}^{\infty}(k-\overline{\kappa})f(k)dk > 0$. Similarly,

$$\begin{aligned}\int_{\underline{\kappa}}^{\infty} kf(k)dk &= \overline{\kappa} + \int_{\underline{\kappa}}^{\overline{\kappa}} (k-\overline{\kappa})f(k)dk + \int_{\overline{\kappa}}^{\infty} (k-\overline{\kappa})f(k)dk \\ &\geq \overline{\kappa} + \int_{\underline{\kappa}}^{\overline{\kappa}} (k-\overline{\kappa})f(k)dk \\ &= \overline{\kappa} + [F(\overline{\kappa})-F(\underline{\kappa})]\frac{\int_{\underline{\kappa}}^{\overline{\kappa}}(k-\overline{\kappa})f(k)dk}{F(\overline{\kappa})-F(\underline{\kappa})} \\ &\geq \overline{\kappa} + \frac{\int_{\underline{\kappa}}^{\overline{\kappa}}(k-\overline{\kappa})f(k)dk}{F(\overline{\kappa})-F(\underline{\kappa})} = B,\end{aligned}$$

where the first inequality holds as it removes something weakly positive, while the second inequality holds because $0 \leq [F(\overline{\kappa})-F(\underline{\kappa})] \leq 1$ and $\int_{\overline{\kappa}}^{\infty}(k-\overline{\kappa})f(k)dk < 0$. It follows that:

$$B < \int_{\underline{\kappa}}^{\infty} kf(k)dk < C \implies B < C.$$

Similarly, it is possible to show that $A < B$. It follows that $A < B < C$.

Since:

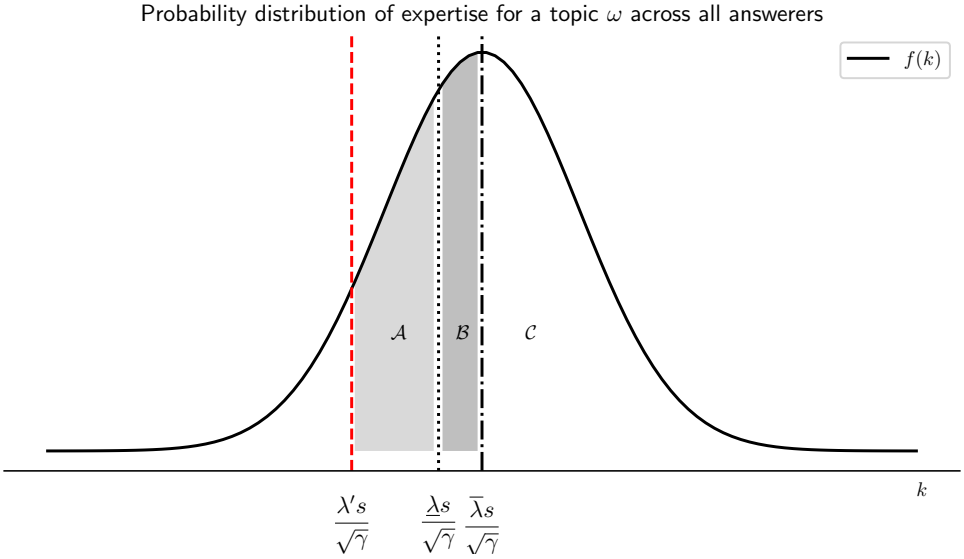
$$\begin{aligned}\mathbb{E}[k_j|\omega, j \in OJ] &= \frac{[F(\overline{\kappa})-F(\underline{\kappa})]B + 2[1-F(\overline{\kappa})]C}{[F(\overline{\kappa})-F(\underline{\kappa})] + 2[1-F(\overline{\kappa})]} \geq C \\ \mathbb{E}[k_j|\omega, j \in NJ] &= \frac{2[F(\underline{\kappa})-F(\kappa')]A + [F(\overline{\kappa})-F(\underline{\kappa})]B}{2[F(\underline{\kappa})-F(\kappa')] + [F(\overline{\kappa})-F(\underline{\kappa})]} \leq C,\end{aligned}$$

then:

$$\mathbb{E}[k_j|\omega, j \in OJ] \geq \mathbb{E}[k_j|\omega, j \in NJ].$$

B.4.3. Case with L Possible Cost Levels for the use of English Consider the general case where there are L possible cost levels (i.e. $\#\Lambda = L$). Define a superscript $l = 1, ..., L$ to index possible cost levels ranked in descending order, such that $\max(\Lambda) = \lambda^1$ and $\min(\Lambda) = \lambda^L$. For the sake of readability, I use the notation κ^l to mean $\frac{s\lambda^l}{\sqrt{\gamma}}$ and define $K \equiv \{\kappa^l\}_{\forall l}$. We then have that:

$$\begin{aligned}\mathbb{E}[k_j|\omega, j \in OJ] &= \sum_{\kappa \in K}\left[\frac{\int_{\kappa}^{\infty} kf(k)dk}{\sum_{\kappa \in K}(1-F(\kappa))}\right] \\ &= \sum_{\kappa \in K}\left[\frac{\int_{\kappa}^{\kappa^1} kf(k)dk}{\sum_{\kappa \in K}(1-F(\kappa))} + \frac{1-F(\kappa^1)}{\sum_{\kappa \in K}(1-F(\kappa))}\frac{\int_{\kappa^1}^{\infty} kf(k)dk}{1-F(\kappa^1)}\right]\end{aligned}$$



Notes. When the cost of language decreases to λ', new answerers with a cost of using English equal to $\underline{\lambda}$ and expertise levels corresponding to the light grey area start to participate. By contrast, the new answerers with a cost of using English equal to $\overline{\lambda}$ have expertise corresponding to both grey areas. Indeed, users in the dark grey area participate in English if they have a low cost of English and do not participate in English if they have a high cost of English. These users may have higher expertise compared to some answerers with a low cost of using English who were already active in English.

Figure EC.1 Probability distribution of expertise across answerers.

$$
\begin{aligned}
&= \sum_{\kappa\in K\setminus\{\kappa^1\}}\left[\frac{\int_{\kappa}^{\kappa^1} kf(k)dk}{\sum_{\kappa\in K}(1-F(\kappa))}\right] + \frac{L(1-F(\kappa^1))}{\sum_{\kappa\in K}(1-F(\kappa))}\frac{\int_{\kappa^1}^{\infty} kf(k)dk}{1-F(\kappa^1)}\\
&= \sum_{\kappa\in K\setminus\{\kappa^1\}}\left[\frac{\int_{\kappa}^{\kappa^2} kf(k)dk}{\sum_{\kappa\in K}(1-F(\kappa))} + \frac{F(\kappa^1)-F(\kappa^2)}{\sum_{\kappa\in K}(1-F(\kappa))}\frac{\int_{\kappa^2}^{\kappa^1} kf(k)dk}{F(\kappa^1)-F(\kappa^2)}\right]\\
&+ \frac{L(1-F(\kappa^1))}{\sum_{\kappa\in K}(1-F(\kappa))}\frac{\int_{\kappa^1}^{\infty} kf(k)dk}{1-F(\kappa^1)}\\
&= \sum_{\kappa\in K\setminus\{\kappa^1,\kappa^2\}}\left[\frac{\int_{\kappa}^{\kappa^2} kf(k)dk}{\sum_{\kappa\in K}(1-F(\kappa))}\right] + \frac{(L-1)(F(\kappa^1)-F(\kappa^2))}{\sum_{\kappa\in K}(1-F(\kappa))}\frac{\int_{\kappa^2}^{\kappa^1} kf(k)dk}{F(\kappa^1)-F(\kappa^2)}\\
&+ \frac{L(1-F(\kappa^1))}{\sum_{\kappa\in K}(1-F(\kappa))}\frac{\int_{\kappa^1}^{\infty} kf(k)dk}{1-F(\kappa^1)}.
\end{aligned}
$$

By iterating the reasoning, we can derive the general statement:

$$
\begin{aligned}
\mathbb{E}\left[k_j|\omega, j\in OJ\right] &= \sum_{l=1,\dots,L-1}\left[\frac{(L-l)(F(\kappa^l)-F(\kappa^{l+1}))}{\sum_{\kappa\in K}(1-F(\kappa))}\frac{\int_{\kappa^{l+1}}^{\kappa^l} kf(k)dk}{F(\kappa^l)-F(\kappa^{l+1})}\right] + \frac{L(1-F(\kappa^1))}{\sum_{\kappa\in K}(1-F(\kappa))}\frac{\int_{\kappa^1}^{\infty} kf(k)dk}{1-F(\kappa^1)}\\
&= \sum_{l=1,\dots,L-1}\left[\frac{(L-l)(F(\kappa^l)-F(\kappa^{l+1}))}{\sum_{\kappa\in K}(1-F(\kappa))}B^l\right] + \frac{L(1-F(\kappa^1))}{\sum_{\kappa\in K}(1-F(\kappa))}C,
\end{aligned}
$$

where $\{B^l\}_{l=1,\dots L-1}$ is the sequence of average expertise values for ranges of k determined by intermediate cost levels of using English. In other words, by letting λ^1 be equal to $\overline{\lambda}$ in the binary cost case and λ^L be equal to $\underline{\lambda}$, the area $\mathcal{B}$ in figure EC.1 is now split in $L-1$ areas defined by the thresholds $\{\kappa^l\}_{\forall l}$. The values

$\{B^l\}_{l=1,...L-1}$ are the average values for each area. Meanwhile, C is the average expertise of *old joiners* that have a cost of using English equal to λ^1.

Similarly, for *new joiners* and $\kappa' \equiv \frac{s\lambda'}{\sqrt{\gamma}}$, we have that:

$$\begin{aligned}
\mathbb{E}\left[k_j|\omega, j \in NJ\right] &= \sum_{\kappa\in K}\left[\frac{\int_{\kappa'}^{\kappa} kf(k)dk}{\sum_{\kappa\in K}(F(\kappa)-F(\kappa'))}\right] \\
&= \sum_{\kappa\in K}\left[\frac{\int_{\kappa^L}^{\kappa} kf(k)dk}{\sum_{\kappa\in K}(F(\kappa)-F(\kappa'))} + \frac{F(\kappa^L)-F(\kappa')}{\sum_{\kappa\in K}(F(\kappa)-F(\kappa'))}\frac{\int_{\kappa'}^{\kappa^L} kf(k)dk}{F(\kappa^L)-F(\kappa')}\right] \\
&= \sum_{\kappa\in K\setminus\{\kappa^L\}}\left[\frac{\int_{\kappa^L}^{\kappa} kf(k)dk}{\sum_{\kappa\in K}(F(\kappa)-F(\kappa'))}\right] + \frac{L(F(\kappa^L)-F(\kappa'))}{\sum_{\kappa\in K}(F(\kappa)-F(\kappa'))}\frac{\int_{\kappa'}^{\kappa^L} kf(k)dk}{F(\kappa^L)-F(\kappa')} \\
&= \sum_{\kappa\in K\setminus\{\kappa^L\}}\left[\frac{\int_{\kappa^{L-1}}^{\kappa} kf(k)dk}{\sum_{\kappa\in K}(F(\kappa)-F(\kappa'))} + \frac{F(\kappa^{L-1})-F(\kappa^L)}{\sum_{\kappa\in K}(F(\kappa)-F(\kappa'))}\frac{\int_{\kappa^L}^{\kappa^{L-1}} kf(k)dk}{\sum_{\kappa\in K}(F(\kappa^{L-1})-F(\kappa^L))}\right] \\
&+ \frac{L(F(\kappa^L)-F(\kappa'))}{\sum_{\kappa\in K}(F(\kappa)-F(\kappa'))}\frac{\int_{\kappa'}^{\kappa^L} kf(k)dk}{F(\kappa^L)-F(\kappa')} \\
&= \sum_{\kappa\in K\setminus\{\kappa^L,\kappa^{L-1}\}}\left[\frac{\int_{\kappa^{L-1}}^{\kappa} kf(k)dk}{\sum_{\kappa\in K}(F(\kappa)-F(\kappa'))}\right] + \frac{(L-1)(F(\kappa^{L-1})-F(\kappa^L))}{\sum_{\kappa\in K}(F(\kappa)-F(\kappa'))}\frac{\int_{\kappa^L}^{\kappa^{L-1}} kf(k)dk}{\sum_{\kappa\in K}(F(\kappa^{L-1})-F(\kappa^L))} \\
&+ \frac{L(F(\kappa^L)-F(\kappa'))}{\sum_{\kappa\in K}(F(\kappa)-F(\kappa'))}\frac{\int_{\kappa'}^{\kappa^L} kf(k)dk}{F(\kappa^L)-F(\kappa')}.
\end{aligned}$$

By iterating the reasoning, we can derive the following general statement:

$$\begin{aligned}
\mathbb{E}\left[k_j|\omega, j \in NJ\right] &= \sum_{l=1,...,L-1}\left[\frac{l(F(\kappa^l)-F(\kappa^{l+1}))}{\sum_{\kappa\in K}(F(\kappa)-F(\kappa'))}\frac{\int_{\kappa^{l+1}}^{\kappa^l} kf(k)dk}{F(\kappa^l)-F(\kappa^{l+1})}\right] + \frac{L(F(\kappa^L)-F(\kappa'))}{\sum_{\kappa\in K}(F(\kappa)-F(\kappa'))}\frac{\int_{\kappa'}^{\kappa^L} kf(k)dk}{F(\kappa^L)-F(\kappa')} \\
&= \sum_{l=1,...,L-1}\left[\frac{l(F(\kappa^l)-F(\kappa^{l+1}))}{\sum_{\kappa\in K}(F(\kappa)-F(\kappa'))}B^l\right] + \frac{L(F(\kappa^L)-F(\kappa'))}{\sum_{\kappa\in K}(F(\kappa)-F(\kappa'))}A,
\end{aligned}$$

where A is the average expertise of *new joiners* with a cost of using English equal to κ^L.

Using the same reasoning as in the binary cost case, it is possible to show that $A < B^{L-1} < ... < B^1 < C$. In addition, as was the case in the binary cost case, it can be noticed that the average expertise of *old joiners* is the weighted average of $\{B^l\}_{l=1,...L-1}$ and A, while the average expertise for *new joiners* is the weighted average of $\{B^l\}_{l=1,...L-1}$ and C. Since the weights are weakly larger on larger values for the *old joiners* and weakly larger on smaller values for the *new joiners*, it follows that:

$$\mathbb{E}\left[k_j|\omega, j \in OJ\right] \geq \mathbb{E}\left[k_j|\omega, j \in NJ\right],$$

as was the case in the binary cost case.

B.5. Remark on the Endogenous Self-Selection of Questioners Across Sites

The theoretical predictions discussed in sections 4.3.1 and 4.3.2 rely on the assumption that the questions published on Stack Overflow are comparable across sites and the language used is the only substantial difference. This implies that Alice expertise k_A and incentives γ are fixed across English and Spanish questions.

This assumption may not hold in reality. Indeed, the choice of publishing a question may depend on site-specific characteristics, leading questioners to self-select on certain sites. A leading factor is the number of

participants, which is smaller on non-English sites and affects the chances of receiving an answer and the speed of answer arrival. Another factor is the availability of expertise in specific topics. Users who are native Spanish speakers may, for instance, be more expert in specific tasks than other users, attracting questioners with questions on those tasks to the Spanish site.

This non-random distribution of questioners and questions across sites may cause systematically different expertise (k_A) and degree of incentives (γ) for the answerer. For instance, if the cultural pool of participants on the Spanish site is more homogeneous and Alice perceives them as culturally closer to her, she may care more for them to solve their problems. In this case, on average, Alice would have a higher γ on the Spanish site than on the English one.[48] At the same time, if the questioners' endogenous choice of websites depends on culture-specific skills, the topics of questions on the Spanish website may be more familiar to Alice than the topics of questions published in English. Her expertise may then be systematically higher on the Spanish website than on the English one.

The data do not allow for testing these hypotheses directly. In the empirical sections, I discuss the implications for identification and develop robustness tests.

C. Details on Estimation

C.1. Details on Identification and Possible Violations of the Parallel Trend Assumption for the Estimation of the *Old Joiners*' Treatment Effect

Consider the example for which Alice is a native Spanish speaker.[49] Let λ_A^{ENG} be her cost of communicating in English when English is her only available language. In addition, let λ_A^{SPA} be her cost of communicating in Spanish. For an answer i published by Alice on the Spanish website at the time t, the treatment effect is given by:

$$\mathbb{E}\left[\Phi_{Ait}\left(\lambda_A^{SPA}\right) - \Phi_{Ait}\left(\lambda_A^{ENG}\right)\right], \tag{EC.2}$$

where $\Phi_{Ait}(\lambda_A^{ENG})$ is the counterfactual quality choice that Alice would have made if she had answered the same question without the option of using Spanish (i.e. the *potential outcome*). While $\Phi_{Ait}(\lambda_A^{SPA})$ is observed, $\Phi_{Ait}(\lambda_A^{ENG})$ is not. The identification of the treatment effect requires the parallel trend assumption and the assumption of no anticipation to hold. More formally, let j index the answers' authors (i.e. Alice, or any other user providing answers) and $\Phi_{Ai(j)t}$ be the quality of the answer i at the time t made by user j. The parallel trend assumption states that $\mathbb{E}\left[\Phi_{Ai(j)t}\left(\lambda_A^{ENG}\right)\right] = \alpha_j + \delta_t$. It implies that, in the absence of the treatment, Alice's quality choice is determined by a time-invariant component specific to her (α_j) and a time-specific

[48] The literature has shown that in various environments, people are more willing to interact with culturally closer individuals because of empathy, trust, and other reasons. For instance, Grinblatt and Keloharju (2001) show that geographical proximity and language drive investors' decisions, while Burtch et al. (2014) show that cultural proximity affects borrowing choices on a crowdfunding platform. In addition, Lyons (2017) shows that teams with the same nationality are more productive, while BenYishay and Mobarak (2019) find that group identity affects communication effectiveness. Ginsburgh and Weber (2020) provide an extensive review of research on how language affects behaviour and economic interactions.

[49] For most of the paper, I use Borusyak et al. (2022) as the main reference for the econometric modelling. While I discuss the empirical strategy and identification using Spanish as the non-English language, the discussion extends to other languages.

component fixed across users (δ_t). In other words, $\mathbb{E}\left[\Phi_{Ai(j)t}\left(\lambda_A^{ENG}\right) - \Phi_{Ai(j)t'}\left(\lambda_A^{ENG}\right)\right]$ is the same across all users (i.e. $\forall j$) for all periods t and t' (whenever $i(j)t$ and $i(j)t'$ are observed). The no-anticipation assumption imposes that when the Spanish website is not yet available, Alice's quality choice is not affected by the fact that it will be.

In the setting of this paper, the no-anticipation assumption is naturally satisfied. In the absence of the Spanish website, Alice can only write in English and her cost of language is not affected even if she anticipates the arrival of the Spanish website. The satisfaction of the parallel trend assumption is more challenging. Indeed, as the theoretical framework shows, several factors affecting the quality choice are answer-specific and are not necessarily fixed across users and time. These include systematic differences across sites on questions' quality, community size, and the kind of topics discussed.

First of all, the question's quality may depend on the questioner's cost of language. Assuming that the questioners writing on the Spanish site are native Spanish speakers, there is the possibility that their questions' average quality in Spanish is higher than on the English site. The theoretical framework would then suggest that Alice's quality choice is significantly higher on the Spanish site independently of changes in her cost of language. To address this problem, I include the question's quality (*QQuality*) in the regression as a control variable.

A second issue that may violate the assumption is that the pool of questioners may be structurally different across websites, as questioners choose where to participate based on site characteristics.[50] This may lead to the following two main confounding effects.

1. Alice's parameter γ may be systematically higher on the Spanish site.

γ represents the degree to which Alice cares for the questioner to solve his problem. If γ is higher on the Spanish site, Alice will choose higher quality in Spanish than in English independently of changes in her cost of language.[51] To address this issue, I exploit variation on the English site over Alice's ability to identify the questioner as a culturally close individual. Indeed, Alice can observe the profile page of the questioner. Since users are free to decide whether to include informative items on their profile page, including their name, place of residence, and profile picture, Alice may or may not learn the cultural identity of the questioner. Using the questioners' profile page data, I construct two proxies of cultural proximity. The first proxy takes a value equal to 1 if the questioner is identified as having the same native language as the answerer, and 0 otherwise. For instance, the first proxy takes a value equal to 1 if the questioner lives in Spain's capital or displays a Spanish name. The second proxy takes a value equal to 1 if the questioner, besides sharing the same native language, displays a manually uploaded profile picture. The latter variable relies on the assumption that the personalised profile picture contains traits supposedly relatable to the answerer if they share the same native language. The analysis includes these proxies as control variables named *empathy*.[52]

[50] Section B.5 in the appendix discusses this issue.

[51] The first derivative of the answerer's best response quality is shown in section B.3 in the appendix.

[52] Section D.3 in the appendix provides details on the construction of these variables. Note that these variables are not available for the *never-treated* users as it is impossible to guess their native language.

2. Alice's expertise (k_A) may be systematically higher on the Spanish site.

This may happen if the topics addressed on the Spanish site are systematically different and more familiar to Alice than the topics discussed on the English site.[53] To address this issue, I execute a test that compares the words in the questions' titles across sites and identifies those that are site-specific. This procedure allows us to exclude answers that may be addressing site-specific topics and check if the results depend on those.[54] A third issue that may harm identification is the difference in community size across websites. As discussed, while this may have an indirect impact through the self-selection of questioners, it can also directly impact the degree of competition. A larger number of competing answerers for the same question can affect the quality of Alice's answer in several ways and in potentially opposite directions. The quality may decrease if the higher competition makes Alice feel more rushed or increase if the competition creates pressure to produce the best answer. In addition, Alice's answer may just complement already existing answers, adding little information. The proxy for quality would be low in this case since it does not incorporate complementarities between answers.[55]

This issue cannot be explained through the model as the model abstracts from strategic behaviour that depends on other answerers' strategies and network effects. To avoid mismeasurement due to this issue, I include measures of competition and active community size as control variables. More precisely, for each answer i in the sample, let $q(i)$ be the question it addresses. The set of control variables referred to as *Competition* includes 1) the number of published answers for question $q(i)$ (including i) and 2) the number of views received by $q(i)$.

D. Details about the Data and the Measures

D.1. Quality Measure based on Code Snippets

Figure EC.2 provides an example of how the quality measure is computed based on code snippets.

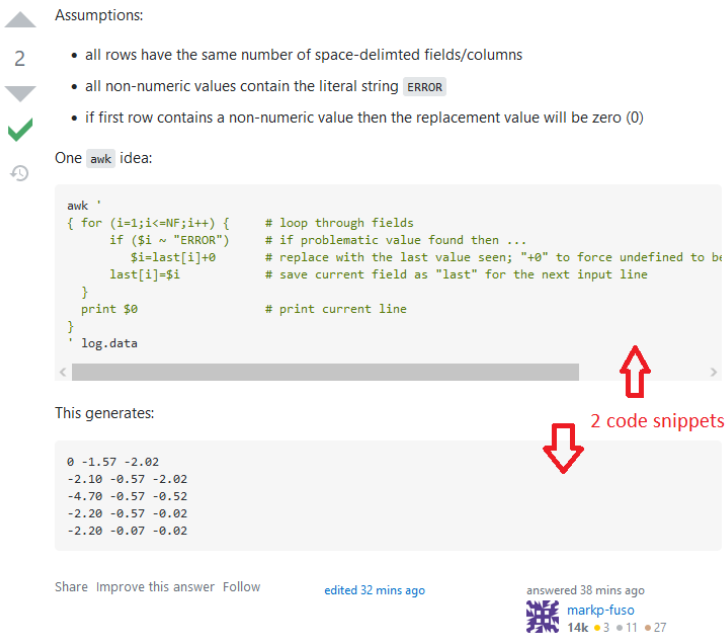
D.2. Correlation between the Proxy for Quality and *Likes*

Table EC.2 reports OLS estimates at the answer level to show the correlation between the variables used as a proxy for quality and the *likes* received. More precisely, the dependent variable in the linear model is the score equal to the number of *likes* minus the number of *dislikes* that the answer received at the time the data were retrieved. The explanatory variables are either dummies equal to 1 if the answer has a corresponding number of pieces of code in the text (columns 1–4) or a dummy equal to 1 if the answer is *accepted* (column 5). Columns 1 and 5 report estimates that use the whole sample. These columns show

[53] Systematic differences in the topics discussed can also be problematic as they may induce mechanical differences on the proxy measure of quality across sites. For example, some topics may not require pieces of code in the answers. This possibility is discussed in section D.2 in the appendix, while robustness analysis is in section E.3 in the appendix. Note that, in this case, there is no prior intuition on the direction of the potential bias.

[54] Details of the procedure and the robustness checks are in sections D.4 and E.2 in the appendix, respectively.

[55] For a different question-and-answer site, Wang et al. (2014) and the references therein find that the number of answers in the thread affects the probability that the questioner solves his problem. While they evaluate the impact on outcomes at the thread level, their evidence suggests a relationship between the answers' quality and the number of answers to the question. In a different context, Jin (2022) shows that competition in information provision with rewards on relative accuracy induces incentives for accuracy but also incentives to differentiate from competitors.



Notes. In this example, there are two snippets of code, as identified by the red arrows. The proxy for quality would then be equal to 2.

Figure EC.2 Example of an answer in Stack Overflow.

that, overall, answers with more pieces of code and *accepted* answers have a higher score. Columns 2 and 3 run the same specification as column 1 but limit the sample to only English and non-English answers, respectively. Column 4 replicates column 3 but excludes answers with no pieces of code. These estimates indicate that while the positive correlation between the number of pieces of code and the score is confirmed in the sample from the English site, it is not in the sample from the non-English site. This is caused by the answers without code. Indeed, excluding those, the positive correlation is re-established.

One possible explanation for these results is that some questions on non-English sites may not require pieces of code in the answer. In that case, answers receive a positive response from the community even if they do not include code. Section E.3 provides robustness analysis to address this issue.

D.3. Construction of Variables to Capture Empathy

***Same-native-language* variable**. For each English answer in the sample, I retrieve the corresponding question and the profile page of the questioner asking that question. I then infer the language of the questioner using two sources: the location and the name displayed.

To infer the language from the location information, I first use the *geopy* Python package to retrieve a standardised format of the location. This process is necessary as users are free to fill the location field in their profile page as they wish. The algorithm attempts to identify the location provided and outputs the name of the country. I then use data from the web to map each country to the main language used.[56]

[56] The data are retrieved from the website https://www.internetworldstats.com/languages.htm, where the main language is the first one in the list provided for each country.

	(1) All sample	(2) Only English	(3) non-English	(4) non-English excl. zero	(5) Score
Number of pieces of code:					
0	0	0	0		
	(.)	(.)	(.)		
1	0.0731	0.211*	-0.914***	0	
	(0.0816)	(0.0953)	(0.0694)	(.)	
[2,3]	0.537***	0.728***	-0.617***	0.297***	
	(0.0794)	(0.0939)	(0.0632)	(0.0570)	
[4,5]	0.903***	1.169***	-0.386***	0.528***	
	(0.0963)	(0.117)	(0.0695)	(0.0634)	
[6,284]	2.046***	2.441***	0.585***	1.498***	
	(0.0885)	(0.109)	(0.0626)	(0.0565)	
Answer is *accepted*:					
No					0
					(.)
Yes					2.330***
					(0.0548)
Constant	2.401***	2.300***	3.126***	2.213***	2.117***
	(0.0602)	(0.0703)	(0.0512)	(0.0448)	(0.0343)
Observations	323850	266568	57282	49440	323850

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

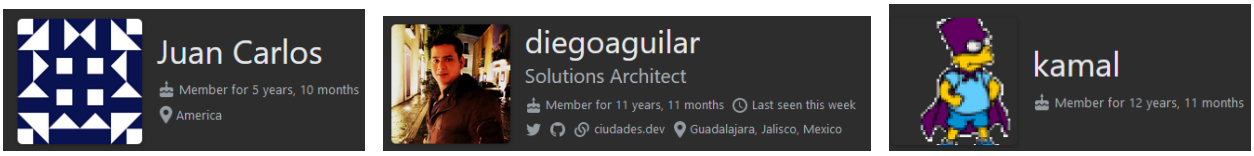
Notes. OLS estimates at the answer level. The dependent variable is the *score* of the answer (i.e. the number of *likes* received minus the number of *dislikes*). In columns 1–4, the explanatory variables are dummies with a value of 1 if the number of pieces of code appearing in the answer falls in a specific range. Estimates show that answers with more pieces of code tend to have a higher *score*. In column 5, the explanatory variable is a dummy equal to 1 if the answer is *accepted* by the questioner as the solution to the problem. The estimates show that answers *accepted* by the questioner tend to have a higher *score*.

Table EC.2 Correlation between different proxies of quality

To infer the language from the name, I proceed as follows. First, I only select users' displayed names that are composed of at least two words, starting with an upper case letter and then at least one lower case letter. I then infer the nationality and the language using the NamePrism API (Ye et al. 2017), which was kindly made available to me by Prof. Steven Skiena and co-authors.[57]

For answers published in a non-English language, I assume that the questioner's native language is the one used on the site. I also assume that the native language of the answerers is the language used on the non-English site where they contribute. This assumption does not help to assign the native language to *never treated* users, for whom I am unable to compute this proxy.

[57] For more details, see https://name-prism.com/api.



Notes. The algorithms used in the paper identify the left and centre users as Spanish native speakers using their name and location, respectively. The left user uses a default avatar, while the centre and right users use personalised images.

Figure EC.3 Examples of user profile page headlines extracted from the English site.

For a given answer, the *same-native-language* proxy is then equal to 1 if 1) the answer is on a non-English site, 2) the questioner's country-based language is the same as the answerer's, or 3) the questioner's name-based language is the same as the answerer's. For all other observations, the proxy is equal to 0.

***Manual-profile-picture* variable**. Similarly to the earlier discussion, I retrieve the question and profile page of the questioner for each English answer in the sample. The data include the URL of the profile picture. If the URL includes the word *gravatar*, then the user is using the default avatar provided by the platform. Otherwise, the user is using a personalised profile picture.

It is plausible to think that, in general, users upload pictures which either represent a photo of the user or an avatar with features representative of the user's identity. Even if this is not the case, the choice of pictures that users upload may still be correlated to their culture. Under this assumption, the presence of a manually uploaded image can increase empathy in the answerer if she relates with it.

For a given answer, the variable *manual-profile-picture* is a dummy that takes a value equal to 1 if the author of the question addressed by the answer displays a personalised picture **AND** if the *same-native-language* variable is equal to 1, and takes a value equal to 0 otherwise.

Examples. Figure EC.3 provides some examples of user profile page headlines which were extracted from the English site. Consider an answer i written in English by an author who is also active on the Spanish site. If i answers a question written by the author on the left or the author on the centre of figure EC.3, then the variable *same-native-language* would be equal to 1 for answer i. This is because the algorithm has identified those two users as Spanish native speakers from the name and the location, respectively. The variable would be equal to 0 if the question's author were the user on the right. Concerning the value of the *manual-profile-picture* variable for answer i, it is equal to 1 only if the author of the question is the user in the centre. This is because that user is the only user with the same native language and a manually uploaded picture. The user on the left displays the default avatar, while the user on the right displays a manually uploaded image; however, the guessed language is different from the guessed language of the answerer (i.e. the *same-native-language* variable is equal to 0).

D.4. Endogenous Selection of Question's Topics' across Sites

To test for the endogenous selection of the questions' topics across sites, I compare the words used in the questions' titles and test whether they are systematically related to a non-English language rather than English. The rationale for using the titles is that, compared to the questions' bodies, they are shorter and less likely to depend on a specific language from a linguistic perspective. This allows for more reliable comparisons

across languages. Moreover, titles generally contain critical information about the topic and the kind of question. In addition, this method allows us to identify which words may be site-specific, allowing an ex-post interpretation of what may drive the selection.

The procedure I follow builds on the intuition of the tests for selection into treatment, which are sometimes used in the literature on field experiments.[58] This corresponds to a joint orthogonality test for a set of observable characteristics, where the dependent variable captures the categories in which the selection may occur. In practice, my approach relies on a logistic regression at the title level. The dependent variable is a dummy equal to 1 if the title is from a non-English site and 0 if it is from the English site. By contrast, the explanatory variables are word frequencies for each word appearing across titles. The regression is run separately for each non-English language site. This method is inspired by the natural language processing (NLP) literature on text classification (e.g. Zhang and Oles (2001)).

For the sake of clarity, I present the detailed procedure using Spanish as an example; the same process has been applied to the other non-English languages.

First, the method pre-processes the data and constructs the regression matrices with the following steps:

1. For all the English and Spanish answers in the sample, retrieve the titles of the corresponding questions.

2. Translate into English all the titles extracted from the Spanish site. To do this, I use Google translate through the Python package *deep_translator*.[59]

3. Parse, separate, and clean the titles' words. This process includes five steps: 1) the selection of only nouns and adjectives via the tokenisation and tagging of the titles; 2) the exclusion of words with less than two characters and more than 15; 3) the transformation of words to lowercase; 4) the exclusion of words commonly considered not meaningful (so-called *stopwords*); and 5) the use of only word roots, which are identified with the Porter stemmer (e.g. *manually* becomes *manual*).[60]

4. Construct a word frequency matrix (explanatory variables) X and the vector of the dependent variable Y. Let i index the questions' titles, N_{eng} be the number of questions from the English site, and N_{spa} be the number of questions from the Spanish site. In addition, let $j = 1, ..., W$ index the unique words extracted from all English and Spanish titles after the pre-processing steps described above. The matrix X has dimensions $(N_{eng} + N_{spa}, W)$ and cell $[i, j]$ reports the frequency of the word j in the title i. The vector Y has dimensions $(N_{eng} + N_{spa}, 1)$ and the element $[i, 1] = 1$ if the title i is from a Spanish question, and 0 otherwise.

With the constructed dataset, the procedure estimates a logit regression model with *l1* regularisation and the liblinear solver.[61] Using the estimated coefficients, the model predicts the probability that a title belongs to the Spanish site. Formally, the predicted probability that a title belongs to the Spanish site is:

$$\hat{p}_i(Y_i = 1|X_i) = \frac{1}{1 + \exp(-X_i\hat{\boldsymbol{\beta}} - \hat{\beta}_0)}.$$

[58] For instance, see: https://blogs.worldbank.org/impactevaluations/tools-trade-joint-test-orthogonality-when-testing-balance.

[59] Credit to Nidhal Baccouri: https://deep-translator.readthedocs.io/en/latest/index.html.

[60] The first part is carried out with the Gensim software and the *simple_preprocess* function. The *stopwords* list is retrieved from the NLTK Python package. Stemming is carried out with the NLTK package and the Porter stemming algorithm (https://tartarus.org/martin/PorterStemmer/).

[61] See https://www.csie.ntu.edu.tw/ cjlin/liblinear/.

A title i is considered problematic if 1) it appears on the Spanish site and 2) the predicted probability that it belongs to the Spanish site is significantly greater than the prior probability. Formally:

$$i \quad \text{is problematic} \iff (Y_i = 1) \quad \text{and} \quad \hat{p}_i^{lb}(Y_i = 1|X_i) > \frac{N_{spa}}{N_{eng} + N_{spa}},$$

where $\hat{p}_i^{lb}(Y_i = 1|X_i)$ is the lower bound of a 95% confidence interval around $\hat{p}_i(Y_i = 1|X_i)$ which is computed via bootstrapping. In other words, a Spanish title is problematic if the correct prediction of its site from the logit classifier is significantly better than the prediction from a random classifier.

Table EC.3 reports the share of questions with problematic titles, conditional on the question's answer(s) being part of the baseline sample. It shows that between 30% and 50% of questions answered in the baseline sample may relate to topics significantly different from those discussed in English. While these numbers are significant, they are upper bounds. Indeed, the prediction exercise that the procedure follows is in-sample. Alternative approaches that test the prediction power of word frequencies out-of-sample would have worse performance. In addition, an assessment of the logit classifier via measures of precision and recall shows a relatively low performance, as shown in figure EC.4. The dots in figure EC.4 represent the so-called precision-recall curve for the logit classifier. A performant classifier would produce a curve close to the top-right corner of the graph. By contrast, a random classifier would produce a horizontal curve at the level of the prior probability (the intermittent line in the graphs).[62] From the graphs, it is possible to infer that, while the model predictions are better than those of a random classifier, it is not achieving good performance. The curve becomes substantially comparable to the curve produced by a random classifier when the problematic questions are removed from the sample. The black star and the thick cross identify the values of precision and recall if we set that:

$$\hat{Y}_i = \begin{cases} 1 & \text{if} \quad \hat{p}_i(Y_i = 1|X_i) > \frac{N_{spa}}{N_{eng}+N_{spa}} \\ 0 & \text{otherwise} \end{cases} .$$

It is possible to notice that at that threshold, the precision value is low while the recall is relatively high. This is in part explainable by the fact that $N_{spa} << N_{eng}$, so the denominator of the recall measure is relatively small.

D.5. Number of Answers per Question on the Native Language Sites

Table EC.5 reports the share of questions published on the native language sites with a certain number of answers and the share and number of answers relevant to those questions. For instance, the first column specifies that 65.77% of questions have only one answer. There are 167,582 of those answers (i.e. answers to questions with only one reply), or 44.1% of the total sample of answers.

E. Robustness for DiD Analysis

E.1. Removing Answers written in English after Treatment by *Treatment Group* Users

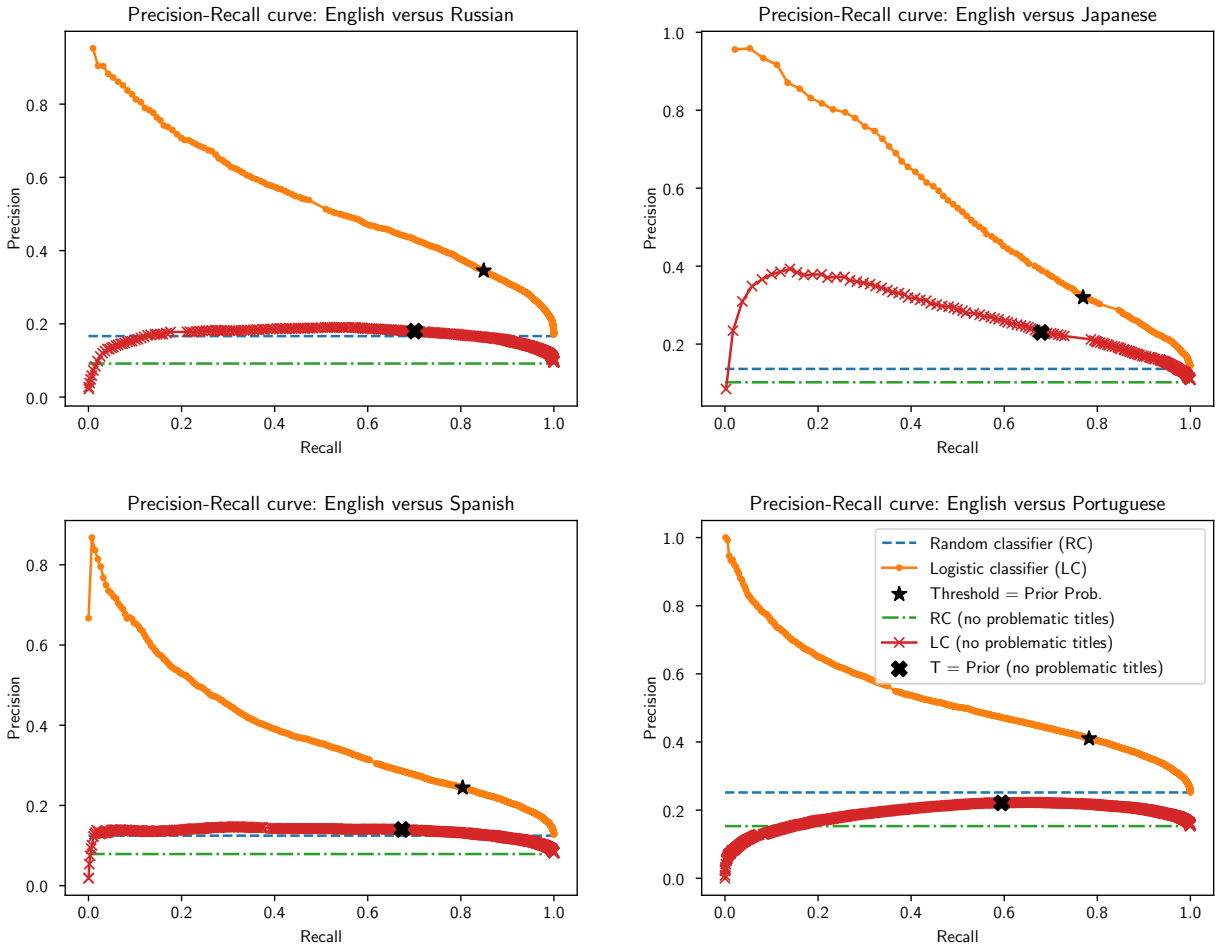
Table EC.6 reports estimates for the main effect of interest after removing *treated* answers written in English by *treatment group* users.

[62] Let a positive prediction be a prediction that a title belongs to the Spanish site. The measure of precision is the ratio between the number of correct positive predictions and the total number of positive predictions. The measure of recall is the ratio between the number of correct positive predictions and the total number of titles belonging to the Spanish site. Since the classifier outputs a probability for the positive prediction, a positive prediction occurs if the predicted probability is higher than an arbitrary threshold. The precision-recall curve computes the measures of precision and recall for a set of threshold values in $(0, 1)$.

Language	Num. Titles	Share problematic
Russian	8293	49.52%
Japanese	3053	28.07%
Spanish	12583	39.9%
Portuguese	25555	46.32%

Notes. Share of titles that the logit classifier has correctly predicted to be in the respective language, doing significantly better than a random classifier. The second column reports the number of observations in each language, which corresponds to the number of questions whose answers appear in the baseline regression sample.

Table EC.3 Share of problematic titles.



Notes. Precision-recall curves for the logit classifier and a random classifier for each language. Each point on the curves corresponds to the classifiers' values of recall and precision for different thresholds. The threshold (t) is an arbitrary value for the probability of the title being assigned to the non-English site. It sets the discriminant of the predicted probability $\hat{p}$ such that if $\hat{p}_i > t$, then $\hat{Y}_i = 1$, and 0 otherwise. The curves are computed for the full sample and the restricted sample that excludes problematic titles. A good classifier should produce a concave curve that bends towards the top-right corner. The figure shows that the logit classifier performs poorly: by removing the problematic titles, it is closely comparable to a random classifier.

Figure EC.4 Precision-recall curves.

Russian	Japanese	Spanish	Portuguese
literatur	rubymin	mercadopago	cpf
uwp	monaca	conda	kotlin
swear	ffmpeg	mercado	monetari
yandex	acquisit	formvalid	nfe
ru	hpack	duda	pagseguro
vk	activerecord	devexpress	bank
phalcon	hoge	androidstudio	made
russian	licens	know	accent
equip	created_at	windowsbuild	demoisel
everyon	casperj	chartj	mandatori
cyril	electron	sii	brazilian
afraid	judgment	navigationview	mount
filestream	bulk	content_main	spservic
memo	countermeasur	lua	sqlsrv
tcpclient	collectionview	exercis	portugues

Notes. List of the 15 words whose frequency vectors have the highest positive coefficient estimates.

Table EC.4 Most language-specific words.

Num. answers per question	1	2	3	4	5+
Share of questions	65.77%	24.2%	7.01%	2.01%	1.02%
Num. Answers in sample	167582	123338	53592	20460	15049
Share of answers in sample	44.1%	32.46%	14.1%	5.38%	3.96%

Notes. Distribution of the number of answers per question for the sample of all non-English questions. The first row provides the share of questions with a given number of answers, while the other rows refer to the answers to those questions, that is, the share of the sample concerned.

Table EC.5 Share of the data by thread length.

	(1) TWFE	(2) TWFE 1	(3) TWFE 2	(4) TWFE 3	(5) BJS	(6) BJS 1	(7) BJS 2	(8) BJS 3
after	1.231**	1.248**	1.254**	1.340**	1.371***	1.411***	1.412***	1.547***
	(0.176)	(0.157)	(0.159)	(0.143)	(0.0334)	(0.0313)	(0.0288)	(0.0167)
Observations	223313	222815	222815	185766	223313	222805	222805	151997
cse	Nat-lang	Nat-lang	Nat-lang	Nat-lang	Nat-lang	Nat-lang	Nat-lang	Nat-lang
Controls								
QQuality	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Competition	No	No	Yes	Yes	No	No	Yes	Yes
Empathy	No	No	No	Yes	No	No	No	Yes

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes. Baseline regression estimates where the dependent variable is the number of pieces of code. The sample exclude *treated* English answers. The estimates correspond to the average treatment effect and correspond to the parameters $\hat{\beta}$ or $\hat{\tau}$ when the specification adopted is the TWFE or the BJS, respectively. The standard errors are clustered (*cse*) at the native language level.

Table EC.6 Old joiners' treatment effect, excluding English answers post-treatment.

E.2. Removing Answers on Site-Specific Topics

Section D.4 shows a procedure to test whether the answers' topics are more likely to appear in a particular language. The procedure estimates the probability that a question belongs to a non-English site using word

frequency vectors, where the words are extracted from the questions' titles. The procedure labels a question as *problematic* if the predicted probability that the question correctly belongs to a non-English site is larger than the probability arising from a random prediction. Answers responding to non-*problematic* questions address topics that, according to the test, are not specifically prevalent in non-English languages.

To ensure that language-specific topics are not driving the results, I estimate the baseline regression from a sample that excludes answers to *problematic* questions. After imposing the sample restriction, the sample is adjusted to ensure that the users in the remaining sample are active both before and after treatment. Table EC.7 reports the results.

	(1) TWFE	(2) TWFE 1	(3) TWFE 2	(4) TWFE 3	(5) BJS	(6) BJS 1	(7) BJS 2	(8) BJS 3
after	0.382^{**}	0.379^{**}	0.380^{**}	0.242^{*}	0.453^{***}	0.465^{***}	0.471^{***}	0.438^{***}
	(0.0654)	(0.0700)	(0.0702)	(0.0543)	(0.0502)	(0.0480)	(0.0453)	(0.0806)
Observations	279181	278402	278402	241353	279181	278333	278333	179516
cse	Nat-lang	Nat-lang	Nat-lang	Nat-lang	Nat-lang	Nat-lang	Nat-lang	Nat-lang
Controls								
QEffort	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Competition	No	No	Yes	Yes	No	No	Yes	Yes
Empathy	No	No	No	Yes	No	No	No	Yes

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

Notes. Baseline regression estimates where the dependent variable is the number of pieces of code. The sample exclude answers that respond to *problematic* questions. The estimates correspond to the average treatment effect and correspond to the parameters $\hat{\beta}$ or $\hat{\tau}$ when the specification adopted is the TWFE or the BJS, respectively. The standard errors are clustered (*cse*) at the native language level.

Table EC.7 Old joiners' treatment effect, excluding language-specific threads.

E.3. Removing Answers with Zero Pieces of Code

Table EC.8 reports regression results comparable to the estimation reported in table 3 after dropping all answers with zero pieces of code and selecting users that, given the remaining answers, were active both before and after treatment.

E.4. Alternative Way to Compute the Average Treatment Effect

The baseline estimation based on Borusyak et al. (2022) computes the treatment effect for each *treated* answer and obtains the average treatment on the treated by averaging the answers' treatment effects with uniform weighting. Since the panel is unbalanced, this approach implies that users who contribute more post-treatment have a larger weight on the final average treatment on the treated.

To obtain an estimate that weighs equally all *treatment group* users, it is possible to average the answers' treatment effects first within each author and then across authors. To apply this approach, it is necessary to modify the third step in the estimation process followed by Borusyak et al. (2022), which is discussed in section 7.2.1. Let $j \in J$ index *treatment group* users and t index time. In addition, define I_j as the set of

	(1) TWFE	(2) TWFE 1	(3) TWFE 2	(4) TWFE 3	(5) BJS	(6) BJS 1	(7) BJS 2	(8) BJS 3
after	0.421*	0.415*	0.415*	0.222	0.670***	0.695***	0.693***	0.773***
	(0.124)	(0.129)	(0.129)	(0.0810)	(0.0351)	(0.0338)	(0.0309)	(0.0521)
Observations	249787	249361	249361	223034	249787	249355	249355	156412
cse	Nat-lang	Nat-lang	Nat-lang	Nat-lang	Nat-lang	Nat-lang	Nat-lang	Nat-lang
Controls								
QQuality	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Competition	No	No	Yes	Yes	No	No	Yes	Yes
Empathy	No	No	No	Yes	No	No	No	Yes

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

Notes. Baseline regression estimates where the dependent variable is the number of pieces of code. The sample excludes answers with zero pieces of code. The estimates correspond to the average treatment effect and correspond to the parameters $\hat{\beta}$ or $\hat{\tau}$ when the specification adopted is the TWFE or the BJS, respectively. The standard errors are clustered (*cse*) at the native language level.

Table EC.8 Old joiners' treatment effect, excluding answers with zero pieces of code.

answers published by the user j after treatment (i.e. when j's native language was already available). If $\hat{\tau}_i$ is the treatment effect for answer i, the average treatment effect is:

$$\text{[Step 3]} \quad \hat{\tau} = \frac{1}{J} \sum_j \left(\frac{1}{\#I_j} \sum_{i \in I_j} \hat{\tau}_i \right).$$

Table EC.9 reports the estimates obtained with this approach and using the baseline sample, while table EC.10 reports the estimates that follow this approach but exclude all *treated* answers in English.[63] It is possible to see that by setting equal weights across users, the effect is much smaller, if significant at all. By focusing only on the non-English answers, the effect is again positive and significant, even if it is smaller.

These results suggest that there is important heterogeneity across users, which is compatible with the results reported in section G.1. Indeed, these results suggest that the treatment effect is larger for users contributing more to the native language site.

E.5. The Role of Zero-Snippets-of-Code Questions for the Complementarity of the Effect with Questions' Quality

Section 7.4.1 tests the hypothesis for which the main effect increases when the question quality is higher. It discusses the possibility that questions with no pieces of code require answers that do not necessarily include pieces of code, and vice-versa. This section provides supporting evidence that the treatment effect increases with higher question quality by addressing this specific issue. First, it shows that by removing questions with no pieces of code, the treatment effect is substantially heterogeneous across levels of question quality. Second, it shows that the probability that the answer is accepted increases with the question's quality.

A possible way to address intrinsic differences between questions with and without code is to remove those without code from the analysis. Table EC.12 reports estimates for the same models used in section 7.4.1. The sample is different as I remove all answers that address a question with zero pieces of code and ensure that

[63] The baseline estimates that use this latter sample are shown in table EC.6.

	(1) BJS	(2) BJS 1	(3) BJS 2	(4) BJS 3
after	-0.0110	0.00171	0.0101	0.0792***
	(0.0255)	(0.0247)	(0.0230)	(0.00957)
Observations	323850	322919	322919	204541
cse	Nat-lang	Nat-lang	Nat-lang	Nat-lang
Controls				
QQuality	No	Yes	Yes	Yes
Competition	No	No	Yes	Yes
Empathy	No	No	No	Yes

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes. Point estimates for the treatment effect using the method employed by Borusyak et al. (2022). It differs from the baseline estimation in how it computes the final average treatment effect. Compared to the baseline, this approach weighs the users equally rather than overweighing users with more contributions post-treatment.

Table EC.9 Old joiners' treatment effect at the author level.

	(1) BJS	(2) BJS 1	(3) BJS 2	(4) BJS 3
after	0.212***	0.227***	0.243***	0.298***
	(0.0253)	(0.0242)	(0.0211)	(0.0234)
Observations	223313	222805	222805	151997
cse	Nat-lang	Nat-lang	Nat-lang	Nat-lang
Controls				
QQuality	No	Yes	Yes	Yes
Competition	No	No	Yes	Yes
Empathy	No	No	No	Yes

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes. Point estimates for the treatment effect using the method employed by Borusyak et al. (2022) and excluding answers written in English post-treatment. Compared to table EC.6, this approach weighs the users equally rather than overweighing users with more contributions post-treatment.

Table EC.10 Old joiners' treatment effect at the author level, excluding English answers post-treatment.

the remaining answerers were active both before and after treatment. Table EC.11 reports the new quality levels for the question. The results show that answer quality increases by 16.8% when the question has only one piece of code and by 32.2% when the question has four or more pieces of code.

A second approach is to use a measure of the answers' quality which does not rely on the amount of code. Table EC.13 reports estimates for the same specifications as in section 7.4.1 but using as dependent variable a dummy equal to 1 if the answer is *accepted* as the solution by the questioner. The table shows that the probability that the questioner *accepts* the answer increases by 13.8% when the question's quality is low and by 22.6% when the question's quality is high.

F. Robustness for the Empirical Analysis of the Study of Selection on Expertise

F.1. Alternative Quality Measures to Study Selection on Expertise

Table EC.14 reports estimate results for the following model (discussed in section 8.2.1):

	Number of snippets of code in the question
Low	{1}
MediumLow	2
MediumHigh	3
High	(3,111]

Table EC.11 Categories for the quality level of the question.

	(1) TWFE	(2) TWFE 2	(3) BJS	(4) BJS 2
Low × after	0.224	0.0644	0.378***	0.460***
	(0.153)	(0.0957)	(0.0414)	(0.0618)
MediumLow × after	0.403*	0.246*	0.589***	0.662***
	(0.114)	(0.0703)	(0.0289)	(0.0452)
MediumHigh × after	0.445*	0.283*	0.658***	0.757***
	(0.111)	(0.0576)	(0.0318)	(0.0497)
High × after	0.576**	0.393**	0.893***	0.883***
	(0.0759)	(0.0323)	(0.0375)	(0.0420)
Observations	245367	218173	245302	152708
cse	Nat-lang	Nat-lang	Nat-lang	Nat-lang
Controls				
QQuality	Yes	Yes	Yes	Yes
Competition	Yes	Yes	Yes	Yes
Empathy	No	Yes	No	Yes

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes. The standard errors are clustered (*cse*) at the native language level, i.e. at the treatment level.

Table EC.12 Estimates by question quality level after dropping observations with zero snippets of code in the question.

$$Quality_i = \alpha_{q(i)} + \delta_{r(i)} + \sum_{g \in G} \beta_g D_{g(j(i))} + \zeta t_{j(i),i} + \varepsilon_i.$$

The dependent variable is a measure of quality for answer i. Different measures use different proxies, including the number of pieces of code included in the answer, a dummy equal to 1 if the answer is *accepted* as the solution to the question, and the number of upvotes received, net of the downvotes. The explanatory variables are, from left to right in the specification, a question fixed effect, an order-of-publication fixed effect, the fixed effect of the authors' group based on their participation on the English site, and the number of days between the author's registration and the publication date.

The results show that, on average, *new joiners* write answers of lower quality than *old joiners*, which is consistent across different quality measures.

	(1) TWFE	(2) TWFE 2	(3) BJS	(4) BJS 2
Low × after	0.00877	-0.00332	0.0764***	0.0497***
	(0.00325)	(0.00278)	(0.00527)	(0.00932)
MediumLow × after	0.0269**	0.0151*	0.0967***	0.0706***
	(0.00456)	(0.00380)	(0.00790)	(0.00485)
MediumHigh × after	0.0307***	0.0189*	0.101***	0.0629***
	(0.00323)	(0.00486)	(0.00650)	(0.0122)
High × after	0.0353**	0.0214	0.109***	0.0813***
	(0.00766)	(0.00857)	(0.00554)	(0.00529)
Observations	322992	285943	322919	204541
cse	Nat-lang	Nat-lang	Nat-lang	Nat-lang
Controls				
QQuality	Yes	Yes	Yes	Yes
Competition	Yes	Yes	Yes	Yes
Empathy	No	Yes	No	Yes

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes. The dependent variable is a dummy equal to 1 if the answer is *accepted* as the solution by the questioner. The standard errors are clustered (*cse*) at the native language level, i.e. at the treatment level.

Table EC.13 Estimates by question quality level.

G. Additional Results

G.1. Heterogeneity on *Old Joiners*' Treatment Effect

Section 7 identified a beneficial effect of introducing multiple languages for *old joiners*. From a managerial perspective, it is relevant to understand whether specific groups of users may be driving the treatment effect.

Proposition 1, which provides the theoretical rationale for the effect, states that the cost of using English is a critical dimension of heterogeneity. Indeed, even if users have the same native language, their cost of using English can differ. By assuming that users have the same cost of using the native language, users with a higher cost of using English face a larger drop in language cost once they use the native language. The comparative statics of the model suggest that the increase in answer quality should be larger for these users.

The researcher does not observe a precise empirical measure for the cost of using English. Nevertheless, this section aims to provide insights into heterogeneity by looking at dimensions that may correlate with it. It investigates heterogeneity based on three considerations: the degree to which the users shifted their contribution to the native language once available, the specific native language, and the intensity of participation in English before the native language became available.

G.1.1. Rate of Adoption of the Native Language Users differentiate themselves in how they shift contributions to the non-English site post-treatment. They may remain active mainly in English, with few contributions in their native language, or prefer to use their native language and eventually abandon the English site. The users' choice may reflect some extra benefit or cost that they receive by participating on

	(1) numCodes	(2) STDnumCodes	(3) Score	(4) STDScore	(5) IsAcceptedAnswer
Registered After Active	-0.932***	-0.218***	-0.832***	-0.244***	-0.0684***
	(0.0497)	(0.0116)	(0.0380)	(0.0111)	(0.00629)
Registered Before Always Active	0	0	0	0	0
	(.)	(.)	(.)	(.)	(.)
Registered Before Active After	-0.647***	-0.151***	-0.495***	-0.145***	-0.0606***
	(0.0622)	(0.0145)	(0.0476)	(0.0139)	(0.00787)
Registered Before Active Before	-1.051***	-0.246***	-0.810***	-0.237***	-0.0687***
	(0.107)	(0.0249)	(0.0816)	(0.0239)	(0.0135)
Registered After Not Active	-1.426***	-0.333***	-1.098***	-0.322***	-0.122***
	(0.0546)	(0.0128)	(0.0418)	(0.0122)	(0.00692)
Registered Before Not Active	-1.856***	-0.433***	-1.197***	-0.351***	-0.147***
	(0.0873)	(0.0204)	(0.0668)	(0.0196)	(0.0110)
Not Registered	-1.707***	-0.399***	-1.274***	-0.373***	-0.144***
	(0.0601)	(0.0140)	(0.0459)	(0.0135)	(0.00760)
DaysFromOldestReg	0.000483***	0.000113***	0.000321***	0.0000939***	0.0000506***
	(0.0000263)	(0.00000615)	(0.0000201)	(0.00000590)	(0.00000333)
Observations	189963	189963	189963	189963	189963
Question Fixed Effects	Yes		Yes		Yes
Order-of-Publication Fixed Effects	Yes		Yes		Yes
Standardized Dep. Var.	No	Yes	No	Yes	No

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

Notes. Regression estimates with different proxies for answer quality. The dependent variable in the specification of column 1 is the number of pieces of code in the answer; in column 2, it is the same variable as in column 1 but standardised. In column 3, it is the number of upvotes minus the downvotes received by the answer; in column 4, it is the same as in column 3 but standardised. In column 5, it is a dummy equal to 1 if the answer has been *accepted*, and 0 otherwise.

Table EC.14 Difference in answer quality between old joiners and new joiners.

the native language site compared to participating in the English one. While many factors can drive this change in utility, variation in the cost of using English could in part explain differences in behaviour.

For a given user, let κ be the number of answers in the native language over the total number of answers written post-treatment. I group users into four categories by their value of κ, using the 25^{th}, 50^{th}, and 75^{th} quantiles as intermediate boundaries. Table EC.15 reports each category's resulting range of values. I then implement the same empirical strategy used in section 7.4 to estimate a category-specific treatment effect. Table EC.16 reports the estimates for the four categories. The parameters correspond to $\{\hat{\beta}_c\}_{\forall c}$ in equation 10 for the TWFE columns and to $\{\hat{\tau}_c\}_{\forall c}$ in equation 11 for the BJS columns. Columns 1–4 report the estimates using the whole sample as described in section 7, while columns 5–8 report the estimates excluding the answers written in English.

Columns 1–4 are more consistent with the definition of treatment used in the baseline regressions. Nevertheless, in this context, these columns are harder to interpret as different categories of users participate on

the English site with different intensities. In particular, the users in the lower categories maintain a stronger presence on the English site after their native language becomes available compared to users in the higher categories. This may confound the interpretation of the results as the users in the lower categories may face a smaller reduction in the cost of language because they mostly keep writing in English after treatment, and not because they have a lower cost of using English. The estimates in columns 5–8 correct for this confounding effect by excluding the answers written in English after treatment.

The estimates show that the effect is largely driven by users who switch to the native language website. This result supports the possibility that users who are less proficient in English have a larger benefit from the introduction of their native language.

	Share of answers not in English in the after-period
Low	(0,0.143]
MediumLow	(0.143,0.425]
MediumHigh	(0.425,0.875]
High	(0.875,1]

Table EC.15 Categories for the rate of adoption of the native language.

	All sample				Sample excludes English Answers Post-Treament			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	TWFE	TWFE 2	BJS	BJS 2	TWFE	TWFE 2	BJS	BJS 2
Low × after	0.0870	0.128	0.145**	0.142	0.483	0.572	0.240***	0.0918
	(0.0972)	(0.104)	(0.0548)	(0.0993)	(0.552)	(0.570)	(0.0498)	(0.0528)
MediumLow × after	0.213	0.105	0.375***	0.200**	1.275**	1.358*	1.224***	0.734***
	(0.108)	(0.107)	(0.0434)	(0.0714)	(0.239)	(0.251)	(0.0345)	(0.0277)
MediumHigh × after	0.648*	0.259	0.469***	0.561***	1.084*	1.168*	0.682***	0.705***
	(0.179)	(0.134)	(0.0451)	(0.0491)	(0.272)	(0.277)	(0.0502)	(0.0456)
High × after	1.462***	0.870**	1.797***	2.088***	1.601**	1.676**	1.859***	2.163***
	(0.143)	(0.147)	(0.0286)	(0.0361)	(0.216)	(0.197)	(0.0279)	(0.0370)
Observations	322992	285943	322919	204541	222815	185766	222805	151997
cse	Nat-lang	Nat-lang	Nat-lang	Nat-lang	Nat-lang	Nat-lang	Nat-lang	Nat-lang
Controls								
QQuality	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Competition	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Empathy	No	Yes	No	Yes	No	Yes	No	Yes

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes. The standard errors are clustered (*cse*) at the native language level, i.e. at the treatment level.

Table EC.16 Estimates of the average treatment effect by the author's share of post-treatment native language answers over the total answers.

G.1.2. Language-Specific Effects The treatment effect may differ across different native languages. For instance, the cost of using English may be lower for users who are native to languages closer to English. To estimate language-specific effects, I use the baseline OLS specification from equation 8 but apply it separately for each language. Since each regression is specific to one of the languages, there is only one treatment date and the standard TWFE method can be used more reliably.

Table EC.17 reports the language-specific estimates for two samples. The first sample (columns 1–4) is the same sample used for the baseline results: it includes all answers published post-treatment by both the *treatment group* and never-treated users. The second set of columns (5–8) reports the treatment effect estimates excluding the answers posted in English by already treated users. While the former sample is more consistent with the previous analysis, the latter sample allows for better comparison across languages since the treatment effect only considers native language answers.

Comparing the two regression sets, the Spanish language stands out, suggesting that Spanish native speakers have different behaviours from the other groups when participating in English post-treatment. The results in columns 5–8 suggest that the effect is slightly larger for Spanish and Portuguese. Since one could think that Spanish and Portuguese are more similar to English, at least to the extent that they have the same alphabet, this result does not support the hypothesis that differences across languages are driven by variations in the cost of using English.[64]

	All sample				Sample excludes English Answers Post-Treament			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Spanish	Portuguese	Russian	Japanese	Spanish	Portuguese	Russian	Japanese
after	0.0831	0.443***	0.298*	0.360***	1.108***	1.177***	0.987***	0.922***
	(0.0799)	(0.0858)	(0.120)	(0.0812)	(0.0898)	(0.0974)	(0.138)	(0.110)
Observations	141917	144402	87589	60231	115766	113326	53116	51754
Controls								
QQuality	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Competition	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Empathy	No	No	No	No	No	No	No	No

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

Table EC.17 Estimates of the average treatment effect for each native language.

G.1.3. Intensity of Participation When it comes to online communities, a small group of users usually contributes the majority of the content. This trend is also evident on Stack Overflow. Table EC.18 reports different category levels for the number of answers published by *treatment group* users in English before treatment. The quartiles of the distribution over the *treatment group* users define each group. The table

[64] Note that a scenario where the cost of using English drives heterogeneity can still be consistent with the results. Let the distribution of the cost of using English across users have large weights on the tails for Russian and Japanese users while being more normally distributed for Spanish and Portuguese users. This suggests that only Russian and Japanese users who are proficient in English participate in pre-treatment as the other users find the cost of using English prohibitively high. However, this pattern does not hold for Spanish and Portuguese users. Consequently, conditionally on the researcher observing their contributions, Russian and Japanese users may experience a smaller reduction in the cost of language once their native language becomes available.

shows that, before treatment, 50% of the users contributed five or fewer answers. By contrast, 25% of the users contributed between 21 and 2,848 answers. The cost of using English may be one reason that drives these participation differences.

Intuitively, users participating more in English pre-treatment may be more proficient in English. The empirical predictions would then suggest that the effect is lower for higher categories of participation. Table EC.19 reports the estimates for each category of contribution intensity. The results suggest that the effect is non-monotonic on the degree answerers were active in English pre-treatment and, overall, do not support the hypothesis.

	Number of answers published in English before treatment
Low	{1,2}
MediumLow	(2,5]
MediumHigh	(5,21]
High	(21,2848]

Table EC.18 Categories for the number of answers published before treatment.

	(1) TWFE	(2) TWFE 2	(3) BJS	(4) BJS 2
Low × after	0.178	-0.187	0.612***	0.395***
	(0.139)	(0.0844)	(0.0533)	(0.0975)
MediumLow × after	0.352	0.0196	0.831***	0.691***
	(0.206)	(0.198)	(0.0342)	(0.0483)
MediumHigh × after	0.262	0.0123	0.648***	0.605***
	(0.130)	(0.0769)	(0.0220)	(0.0306)
High × after	0.393*	0.253*	0.559***	0.615***
	(0.0980)	(0.0586)	(0.0436)	(0.0716)
Observations	322992	285943	322919	204541
cse	Nat-lang	Nat-lang	Nat-lang	Nat-lang
Controls				
QQuality	Yes	Yes	Yes	Yes
Competition	Yes	Yes	Yes	Yes
Empathy	No	Yes	No	Yes

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

Notes. The standard errors are clustered (*cse*) at the native language level, i.e. at the treatment level.

Table EC.19 Estimates of average treatment effect by author's intensity of contribution in English before treatment.

G.2. Externalities on the English Site

The theoretical framework's assumptions do not allow for quality spillovers across sites. Indeed, the choice of quality on the English site does not depend on the choice of quality on the native language site or any other native language site characteristics. There is no straightforward reason why the existence of a native language site may impact the users' quality choices on the English site. Nevertheless, behavioural factors

may create such spillovers if the users consider the native language and English sites complementary or substitutable.

To test for externalities on the English site, I estimate the *old joiners*' treatment effect excluding non-English contributions. In other words, I replicate the baseline analysis in section 7 but use a sample that only includes answers written in English. Table EC.20 reports the estimates for β in equation 8 in the context of TWFE regression and for τ in equation 9 in the context of BJS estimation. The results suggest that the introduction of multiple languages had positive spillovers on the English website even though these were not significant in the preferred specification. The channel of this positive effect remains an open question.

	(1) TWFE	(2) TWFE 1	(3) TWFE 2	(4) TWFE 3	(5) BJS	(6) BJS 1	(7) BJS 2	(8) BJS 3
after	0.0449	0.0512	0.0505	0.0677	0.117*	0.119*	0.125*	0.147
	(0.0506)	(0.0541)	(0.0541)	(0.0804)	(0.0576)	(0.0554)	(0.0530)	(0.0976)
Observations	261771	260950	260950	223901	261771	260887	260887	176268
cse	Nat-lang	Nat-lang	Nat-lang	Nat-lang	Nat-lang	Nat-lang	Nat-lang	Nat-lang
Controls								
QQuality	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Competition	No	No	Yes	Yes	No	No	Yes	Yes
Empathy	No	No	No	Yes	No	No	No	Yes

Standard errors in parentheses
* $p<0.05$, ** $p<0.01$, *** $p<0.001$
Notes. The standard errors are clustered (*cse*) at the native language level, i.e. at the treatment level.

Table EC.20 Estimates of the treatment effect on the quality of English answers.

Acknowledgments

I am extremely grateful to three anonymous referees, the associate editor, and the editor for comments that substantially improved the paper. I would like to thank for helpful discussions and feedback Patrick Bennett, Milo Bianchi, Gordon Burtch, Daniel L. Chen, Jacques Crémer, Daniel Ershov, April Franco, Guido Friebel, Astrid Hopfensitz, Jakub Lonsky, Nicolas de Roos, Marta Troya-Martinez, Shalini Mitra, and Bobby Zhou. I thank for useful comments participants at the University of Liverpool Management School internal seminar, Digital Economy Network workshop, Munich Summer Institute, SIOE, Platform Strategy Research Symposium, and ESEM.