Low-cost Language Learning: A boost to move?*

Evidence from *Duolingo*

Joop Adema

ifo institute & LMU

August 26, – EEA 2024 Rotterdam

*The author thanks Padmaja Kadambi, Tuncay Taghiyev and Lara Bieske for excellent research assistance

Motivation: Language Learning apps

- Mobile internet adopted by more than 4 billion people worldwide
- Many aspects of the internet technologies are relevant to migration: ease of communication, search for information, social media (Adema et al., 2022)
- A popular internet technology is **low-cost Language Learning** through apps such as *Duolingo* (83m monthly active users)
- Language skills are important to migrants' choice of destination and lack thereof limits integration into host societies
- Conventional language learning is expensive $\rightarrow$ Language learning apps reduce language learning costs

$\rightarrow$ did the introduction of low-cost language learning impact language skills, international migration patterns and migrant integration?

This paper

How does the availability of low-cost language learning affect migration aspirations and flows?

- Duolingo enables learning from a *source* to a *target* language for 84 courses introduced in a staggered fashion between 2012 and 2021
- Link rollout to English test scores (TOEFL)
- Construct a dyadic country-level measure of Duolingo exposure depending on course rollout and languages spoken by country and link it to bilateral migration aspirations and flows
- Link rollout to migrants' self-assessed language skills in destination country surveys

Low-cost language learning...

- Improved reading and listening (but not writing and speaking) skills among English test takers
- Increased bilateral migration aspirations, and flows to OECD countries after 3 years
- Improved migrants' language skills in the US, without affecting migrants' self-selection

The state of the literature

Language explains international migration flows:

- Shared languages and linguistic proximity lead to larger bilateral migration flows Belot and Ederveen (2012); Adsera and Pytlikova (2015)

Language knowledge is important for labor market integration of immigrants:

- Relevant language skills increases earnings by 5 to 35 % across studies (Adsera and Pytlikova (2016)
- Governments spend a lot of money on language courses, which improve labor market outcomes Sarvimäki and Hämäläinen (2016); Foged, Hasager and Peri (2022)
- Language course enrollment increases civic engagement and earnings in the US, costs well exceed benefits Heller and Slungaard Momma (2023)

The opportunity of language learning has been shown to affect international migration:

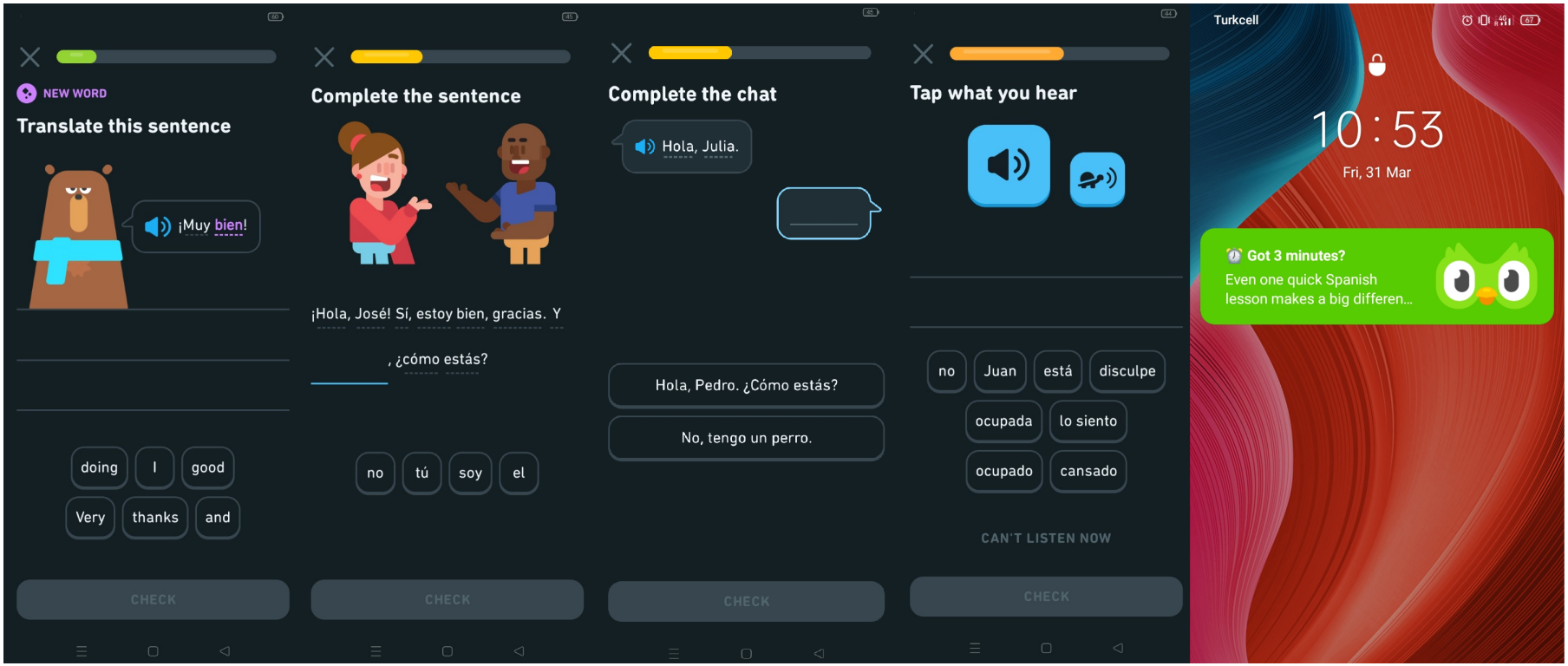
- The presence of German language institutes increase migration to the German-speaking world Huber and Uebelmesser (2023)
- Moreover, migrants have better German skills upon arrival and are better educated. Jaschke and Keita (2021)

Duolingo concerns language learning **before and after departure**, that is **free to use** and **does not offer certification**

Duolingo: an educational technology

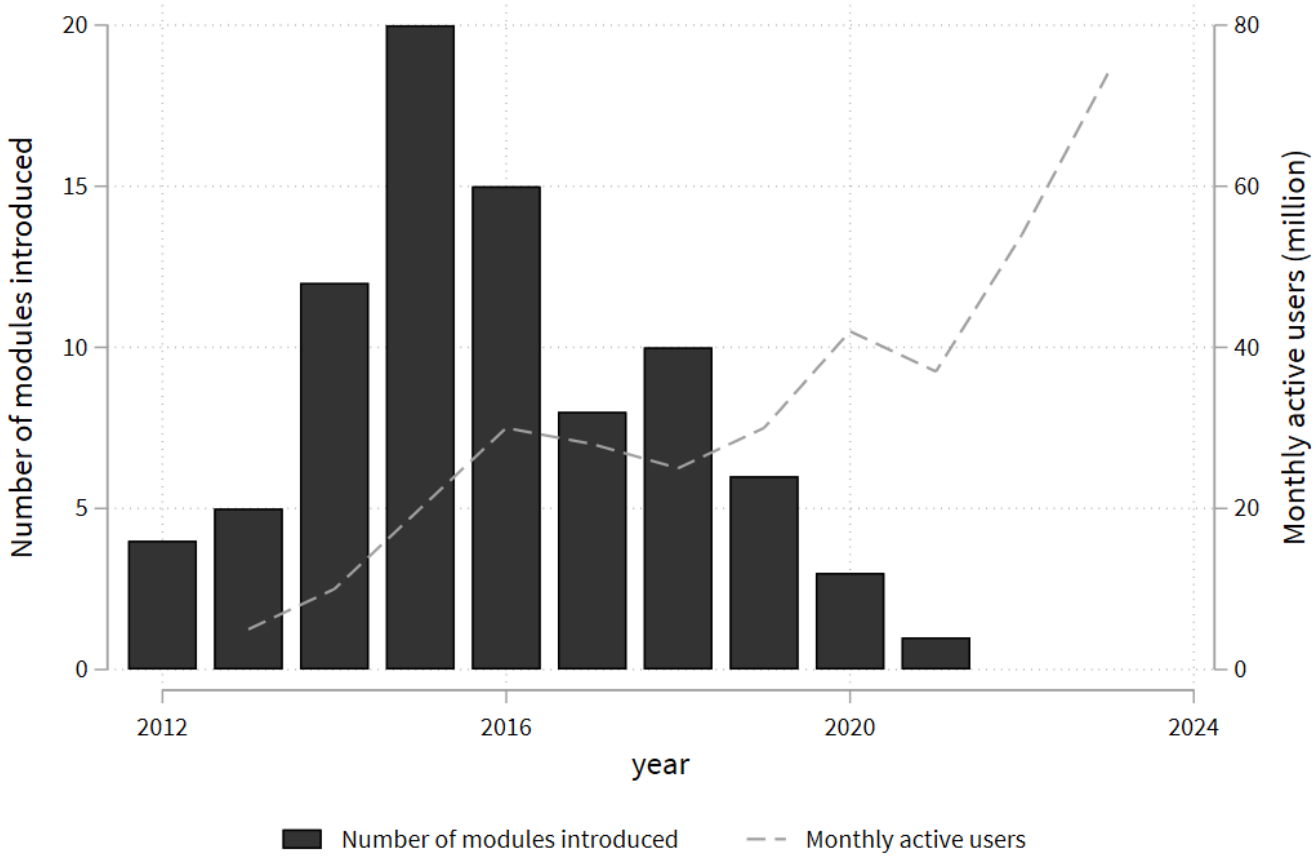
- Established in 2011
- Comprises *courses* that enable one to learn a **target** language from a specific **source** language, e.g. English to Spanish
- Staggered rollout of 84 language learning courses between 2012 and 2021 between 23 distinct source and 30 distinct target languages
- Courses are extensive: 10,000s of sentences; courses reach up to and including CEFR level B2
- Has been found to be effective at beginner levels Vesselinov and Grego (2012); Rachels and Rockinson-Szapkiw (2018); Ersoy (2021)
- Basic version available for free, currently 83 million monthly active users and a 68% market share

Duolingo: an educational technology...and a commitment device

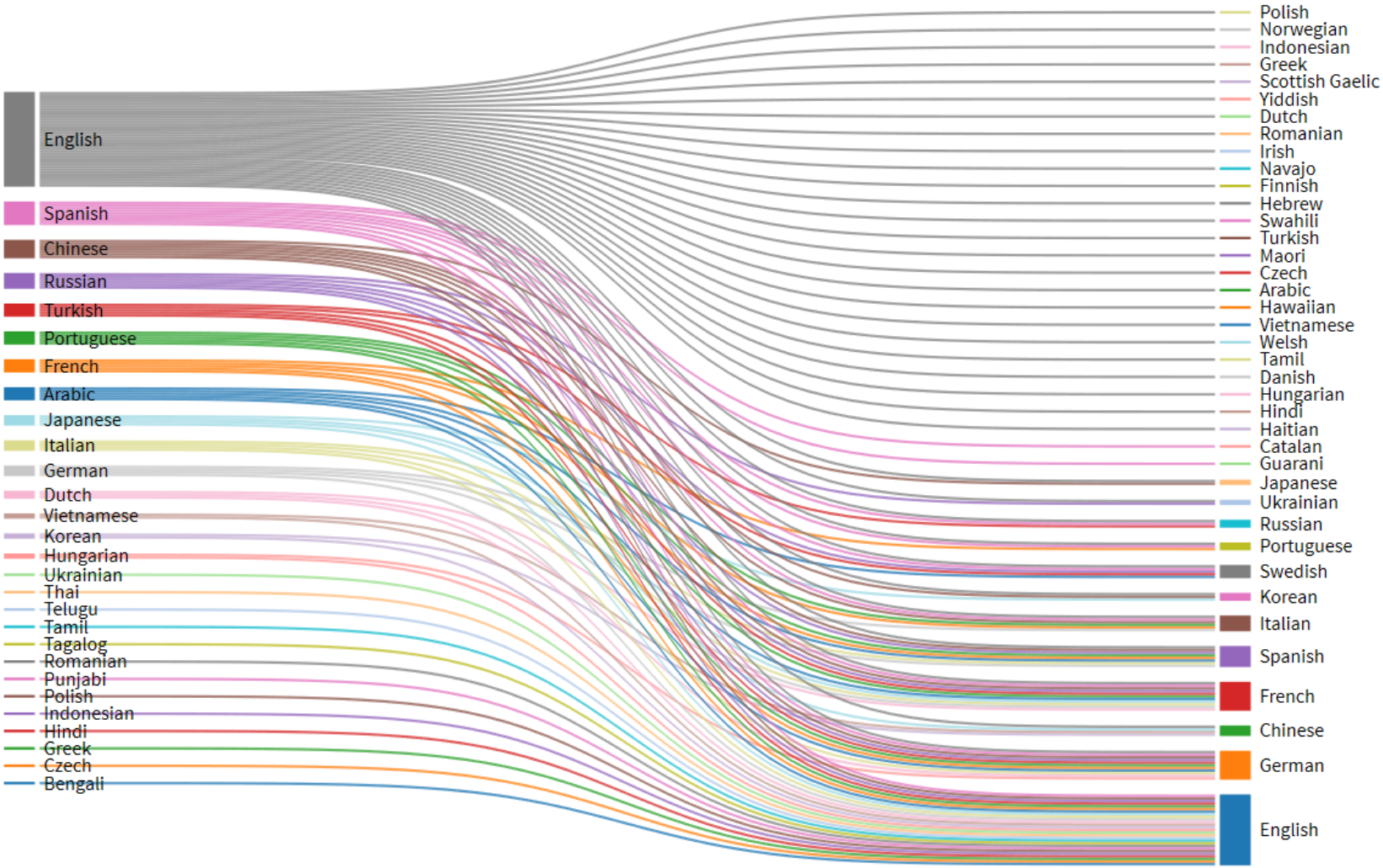


Mostly *passive* elements of language learning

Duolingo: course rollout



Duolingo courses currently available



From model to empirical strategy

Random utility model: $U_{iod} = \ln w_{iod} = \mu_{od} + b_{odT} l_{oT} + \epsilon_{iod}$

$$\ln\left(\frac{\mathbb{P}_{od}}{\mathbb{P}_{oo}}\right) = \mu_{od} - \mu_{oo} + (b_{odT} - b_{ooT}) l^*_{oT}$$

- μ's: wage levels b's: returns to language skills l in target language T
- l^*_{oT} decreasing (increasing) in learning costs (Duolingo availability)
- Sign depends on the relative strength of the domestic and foreign returns
- Assuming language skills are randomly distributed among its population, I proxy b's with the probability that a Duolingo course enables communication between two random individuals: DL^{ST}_{od}
- Requires the share of speakers by language and country

$$DL^{abroad}_{odt} = \max_{S,T} DL^{ST}_{od} \, Duolingo_{STt}$$

$$DL^{domestic}_{oot} = \max_{S,T} DL^{ST}_{oo} \, Duolingo_{STt}$$

Identification and estimation

Based on an RUM model, we estimate a gravity model with three-way FE:

$$\frac{M_{odt}}{M_{oot}} = exp\left[\beta_1 DL_{odt}^{abroad} + \beta_2 DL_{oot}^{domestic} + \psi_{od} + (\phi_{ot}) + \theta_{dt}\right]\eta_{odt}$$

- M_{odt} is the stock of individuals aspiring to migrate from o to d at t
- M_{oot} is the stock of individuals not aspiring to migrate from o at t
- ϕ_{ot}, θ_{dt}, ψ_{od} are three-way fixed effects capturing unobserved heterogeneity at the dyad level, as well at the origin-year and destination-year level
- η_{odt} is a mean-1 idiosyncratic error term
- Estimate model by PPML, β's identify a relative average effect
- We cluster the standard errors at the origin- and destination level

Main identifying assumption: parallel trends in proportions

The outcome in treated dyads would have had the same relative growth as untreated units would Duolingo not have rolled out, net of origin-time and destination-time FE.

- Unlikely that Duolingo courses are rolled out in anticipation of increasing migration aspirations:
  - Duolingo was likely supply-constrained and courses required time to build
  - *Duolingo* not primarily used for migration preparations (but for schooling, work, brain training and family-related reasons)
  - Many courses built by the user community who have no commercial motive
  - Anticipating aspirations is unlikely as it is the precedes migration
- We can test the trends prior to roll-out: Event study around large increases in Duolingo exposure shows no pre-trends
- To alleviate any remaining concerns, for each of the courses we can remove the contribution to DL_{odt} for the origin country with the most source language speakers and for the destination country with the most target language speakers.

Data on Duolingo rollouts and languages spoken

Rollout of *Duolingo* courses

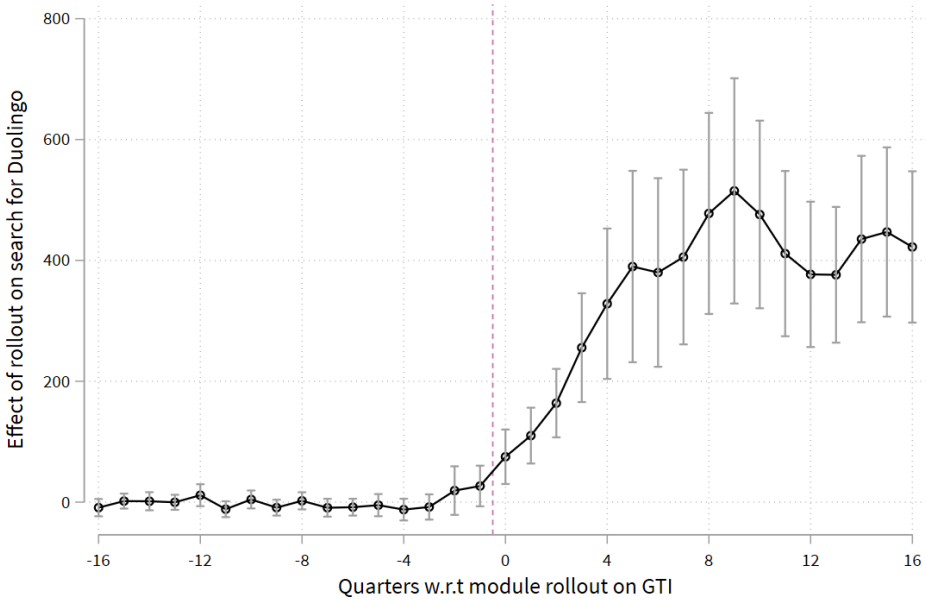
- Timing of course rollout from online sources
- Validated through online search interest (Google Trends)

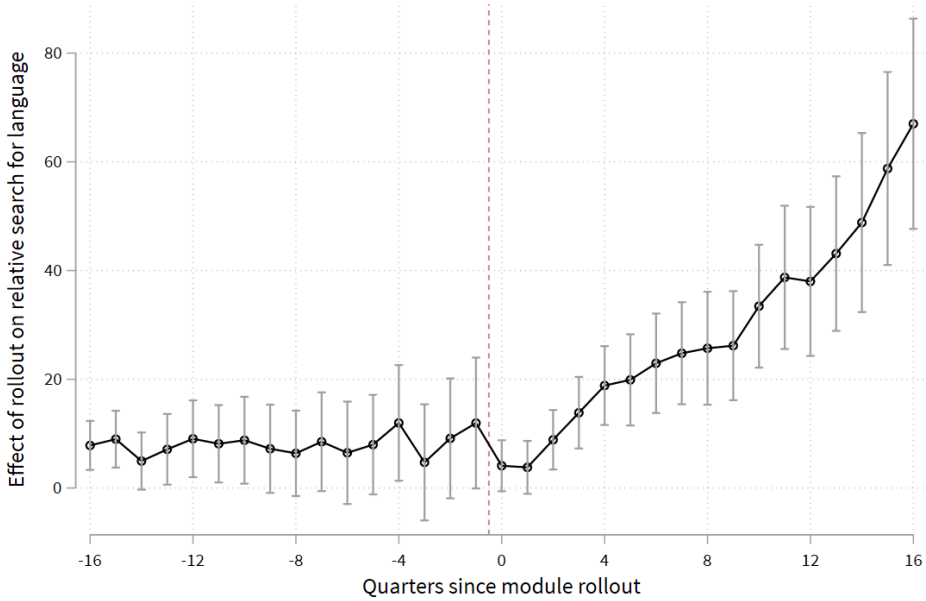
Languages spoken by country:

- Ginsburgh et al. (2017) construct a dataset of languages by country based on Eurobarometer ("do you speak well enough in order to be able to have a conversation?"), Ethnologue and the CIA World Factbook
- We use *spoken* source and target languages rather than official languages as those can be used to acquire a foreign language, or pay off on foreign labor markets (e.g. English in the Netherlands)

Do course rollouts spur search interest in Duolingo and languages?

Event studies of Google Trends Index in Duolingo and languages



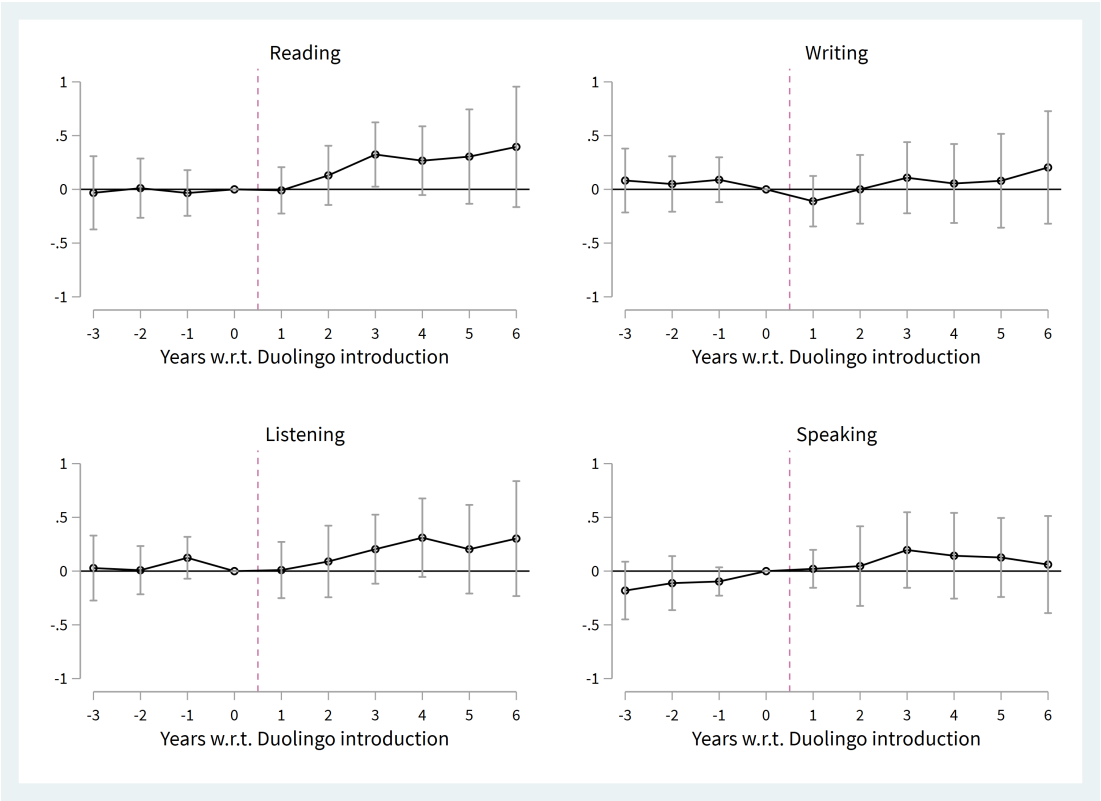


- Interest in *Duolingo*
- First relevant (> 0.5) DL course
- TWFE at origin-time level
- Roll-out dates are meaningful

- Interest in *Target Language*
- First relevant (> 0.5) DL course
- 3WFE at origin-target-month level
- Increase is very gradual

Language skills (2007 - 2022)

- TOEFL: English test taken by more than 2 million people every year
- Average scores by native language for four components
- Estimation by Borusyak-Jaravel-Spiess (2023) estimator



Effect size: about 0.1 S.D.

Data on (migration aspirations)

Gallup World Poll 2007-2023 (GWP): a representative survey of generally around 1000 individuals per year in more than 150 countries, probes migration *aspirations* in the following way:

- Ideally, if you had the opportunity, would you like to move permanently to another country, or would you prefer to continue living in this country? **22%** say yes.
- To which country would you like to move?

We construct the stock of people in country o aspiring to emigrate to country d from the origin country's population and the share of respondents aspiring to emigrate from o to d:

$$M_{odt} = pop_o \times \frac{N_{odt}}{N_{ot}}$$

- Tjaden et al. (2019) show this correlates strongly to *subsequent* migration flows Estimation sample: 153 origin countries, 193 destination countries and 16 years
- OECD *actual bilateral migration* data: 196 origin countries, 36 destination countries and 16 years

The effect of Duolingo on bilateral migration aspirations

	(1)	(2)
DL_{odt}^{abroad}	0.273***	0.400***
	(0.070)	(0.083)
$DL_{odt}^{domestic}$	-0.427**	
	(0.200)	
Observations	123484	123484
Unique origin countries	153	153
Unique destination countries	193	193
Unique dyads	10663	10663
Origin-destination FE	✓	✓
Origin-year FE		✓
Destination-year FE	✓	✓

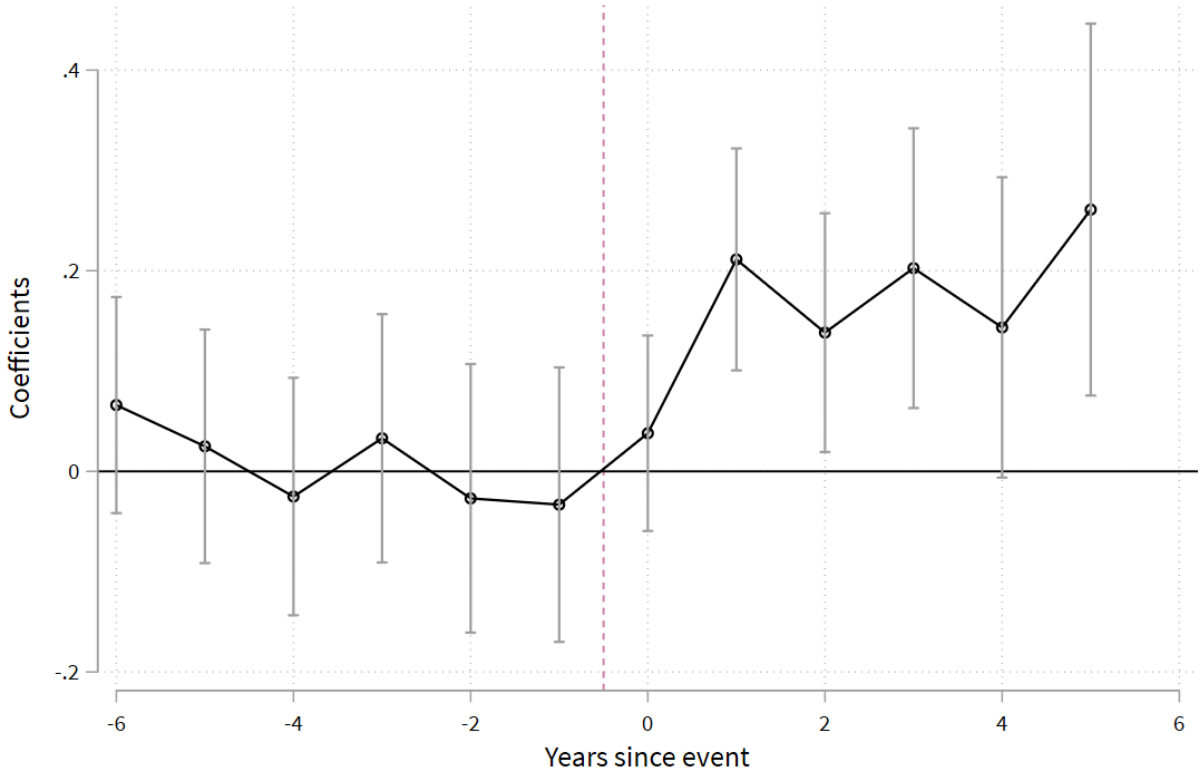
Estimated by PPML. Standard errors are clustered at origin-destination level.

In line with the model, domestic Duolingo exposure *decreases* and foreign Duolingo exposure *increases* the log odds of migration

Event study around large (more than 70 pp) course introductions

We use the Borusyak-Jaravel-Spiess (2021) imputation-based estimator: accounts for three-way fixed effects.

- Treatment: $max(DL_{odt}^{abroad}) > 70$ pp
- Control: $max(DL_{odt}^{abroad}) < 20$ pp



Note: Standard errors clustered at origin and destination level.

The role of English

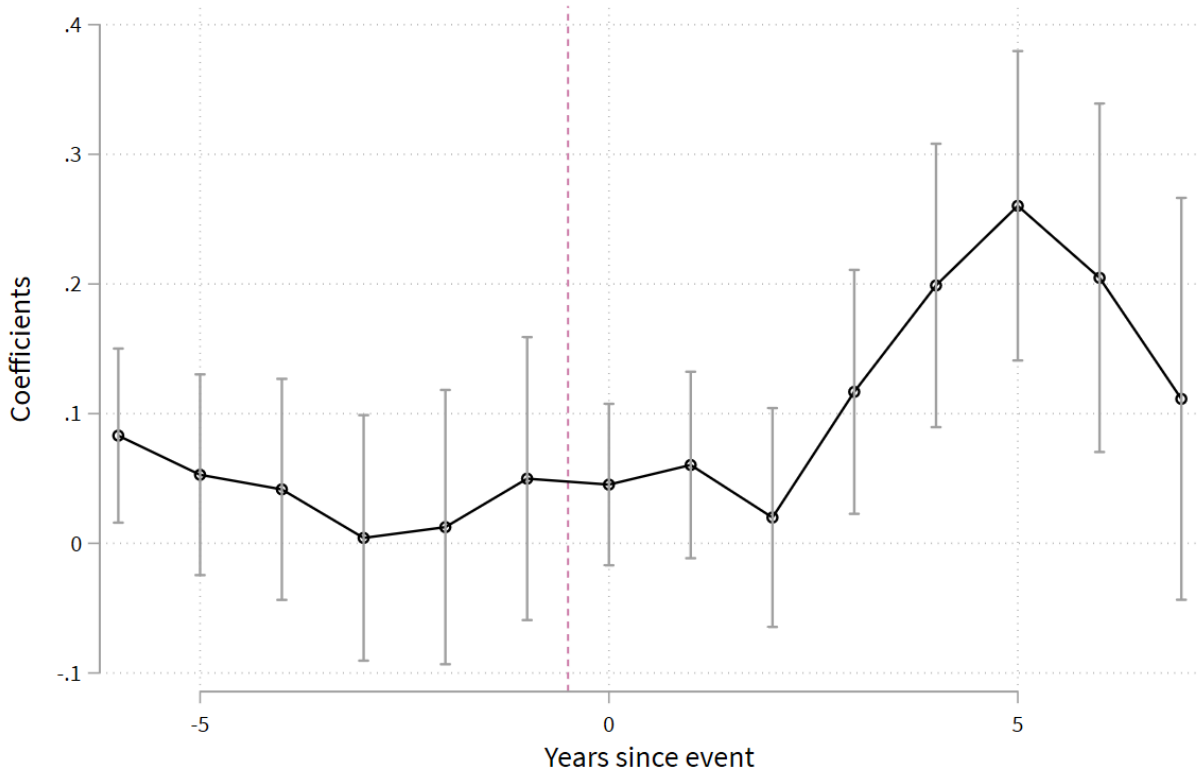
	(1)	(2)	(3)	(4)
$DL_{odt}^{S=EN,abroad}$	0.772***	0.899***		
	(0.205)	(0.208)		
$DL_{odt}^{S!=EN,abroad}$	0.194**	0.295***		
	(0.082)	(0.100)		
$DL_{odt}^{S=EN,domestic}$	-0.248			
	(0.241)			
$DL_{odt}^{S!=EN,domestic}$	-0.434*			
	(0.256)			
$DL_{odt}^{T=EN,abroad}$			0.186**	0.581***
			(0.092)	(0.092)
$DL_{odt}^{T!=EN,abroad}$			0.239**	0.106
			(0.106)	(0.092)
$DL_{odt}^{T=EN,domestic}$			-0.507*	
			(0.285)	
$DL_{odt}^{T!=EN,domestic}$			-0.074	
			(0.221)	
Observations	124878	123484	124878	123484

Effect heterogeneity

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
DL_{odt}^{abroad}	0.347***	0.206**	0.487***	0.400***	0.350***	0.054	0.645***
	(0.079)	(0.086)	(0.101)	(0.083)	(0.097)	(0.233)	(0.137)
× GTI for Duolingo	0.000**						
	(0.000)						
× $3G_{oy}$ (Collins)		0.327**					
		(0.153)					
× $Broadband_{oy}$ (ITU)		-0.002					
		(0.005)					
× Linguistic proximity			-0.527**				
			(0.243)				
× Shared official language				-0.406**			
				(0.198)			
× PPP_{oy}					0.003		
					(0.003)		
× PPP_{dy}						0.007	
						(0.005)	
× Log of migrant stock							-0.027*
							(0.016)
Observations	121385	89933	114404	122461	118102	115367	122161

Actual migration to OECD countries (2007 - 2019)

We use the Borusyak-Jaravel-Spiess (2021) estimator using the bilateral yearly migration flows to 37 OECD countries



→ Actual migration flows start increasing after 3 years

Conclusion and outlook

The availability of low-cost language learning:

- Increases the interest in a foreign languages and TOEFL test takers' reading and listening scores by 0.1 S.D.
- Spurs international migration aspirations, actual migration three years after course introduction and improves the language skills of migrants

Policy relevance:

- Governments spend a lot of money on integration and language courses, but modern technology may help quite a bit
- Even though this is a low-cost technology, it does not seem to make the migrant pool lower-skilled

The next steps:

- Pin down effects on language skills *of the general EU population* (AES)
- Study effects of better language skills in the EU (LFS) and US (ACS); employment, earnings, and occupational choice
- Duolingo has a wealth of data behind the scenes...

Migrants' language skills and integration in the USA

United States has:

- Diverse migrant pool, 22 modules to English, 9.1 million learners of English in the US on Duolingo in 2007
- The American Community Survey (ACS) surveys 3.5 million households yearly (including (undocumented) immigrants and family migrants)
- Questions on demographics, labor market, year of migration, country of birth and self-assessed language skills:
  1. Not at all (0.112)
  2. Not well (0.131)
  3. Well (0.339)
  4. Very well (0.240)
  5. Speaks only English (0.178)

We focus on immigrants who moved to US at ages 18-60 who have immigrated to the USA in or after 2007.

$$y_{itco} = \alpha DL_{co}^{pre} + \beta DLyears_{tco}^{post} + \gamma' \mathbf{X_{itc}} + \delta' \mathbf{Z_{c-1o}} + \phi_{c(t-c)} + \psi_{o(t-c)} + \epsilon_{itco}$$

Language skills

	Speaks EN	Speaks EN at least well	Speaks EN at least very well	Age at im-migration	Female	At least 9th grade (at least 18)	At least some tertiary education (>25)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
DL^{pre}_{oc}	0.008**	0.009	0.007	0.167	0.004	0.009*	-0.008
	(0.003)	(0.007)	(0.006)	(0.132)	(0.004)	(0.005)	(0.005)
$DLyears^{post}_{oc}$	0.003**	0.009***	0.007**	0.116*	-0.002	0.003	0.010*
	(0.001)	(0.002)	(0.003)	(0.066)	(0.002)	(0.002)	(0.005)
Observations	400217	400217	400217	400217	400217	343283	245411
R^2	0.23	0.33	0.30	0.07	0.03	0.21	0.32
Average dependent variable	0.889	0.716	0.477	29.628	0.506	0.902	0.639
Fixed Effects	✓	✓	✓	✓	✓	✓	✓
Controls for age and sex	✓	✓	✓				

Estimated by OLS, Standard errors clustered at the country of origin.

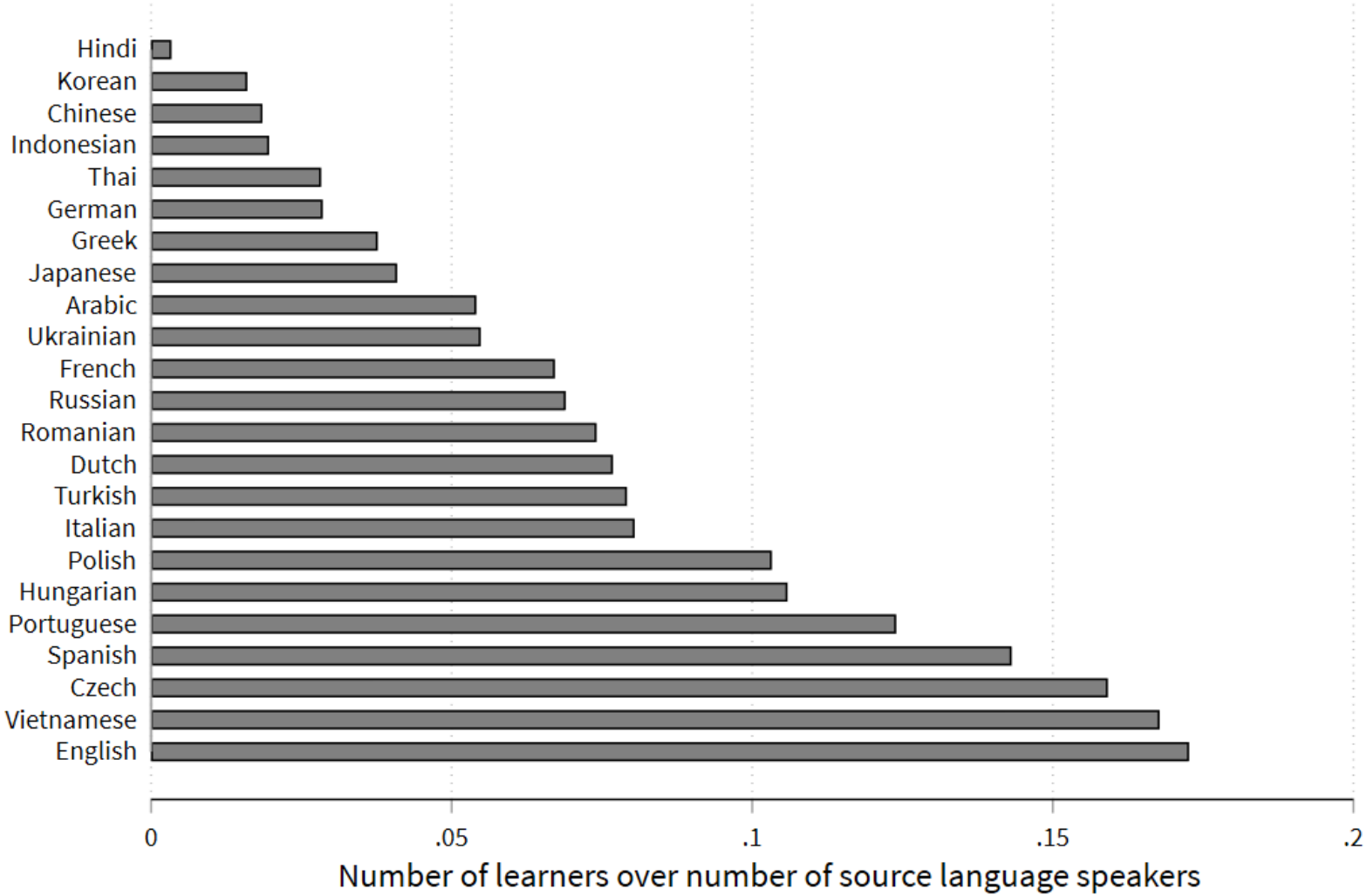
Labor market outcomes

	Employed	Log of Total income earned + 1	Occupational score
	(1)	(2)	(3)
DL^{pre}_{oc}	-0.003	-0.055	-0.108
	(0.005)	(0.052)	(0.163)
$DLyears^{post}_{oc}$	0.001	0.016	0.166
	(0.003)	(0.039)	(0.106)
Observations	400217	379120	400217
R^2	0.25	0.22	0.29
Average dependent variable	0.531	6.026	19.014
Fixed Effects	✓	✓	✓
Controls for age, age2 and sex	✓	✓	✓

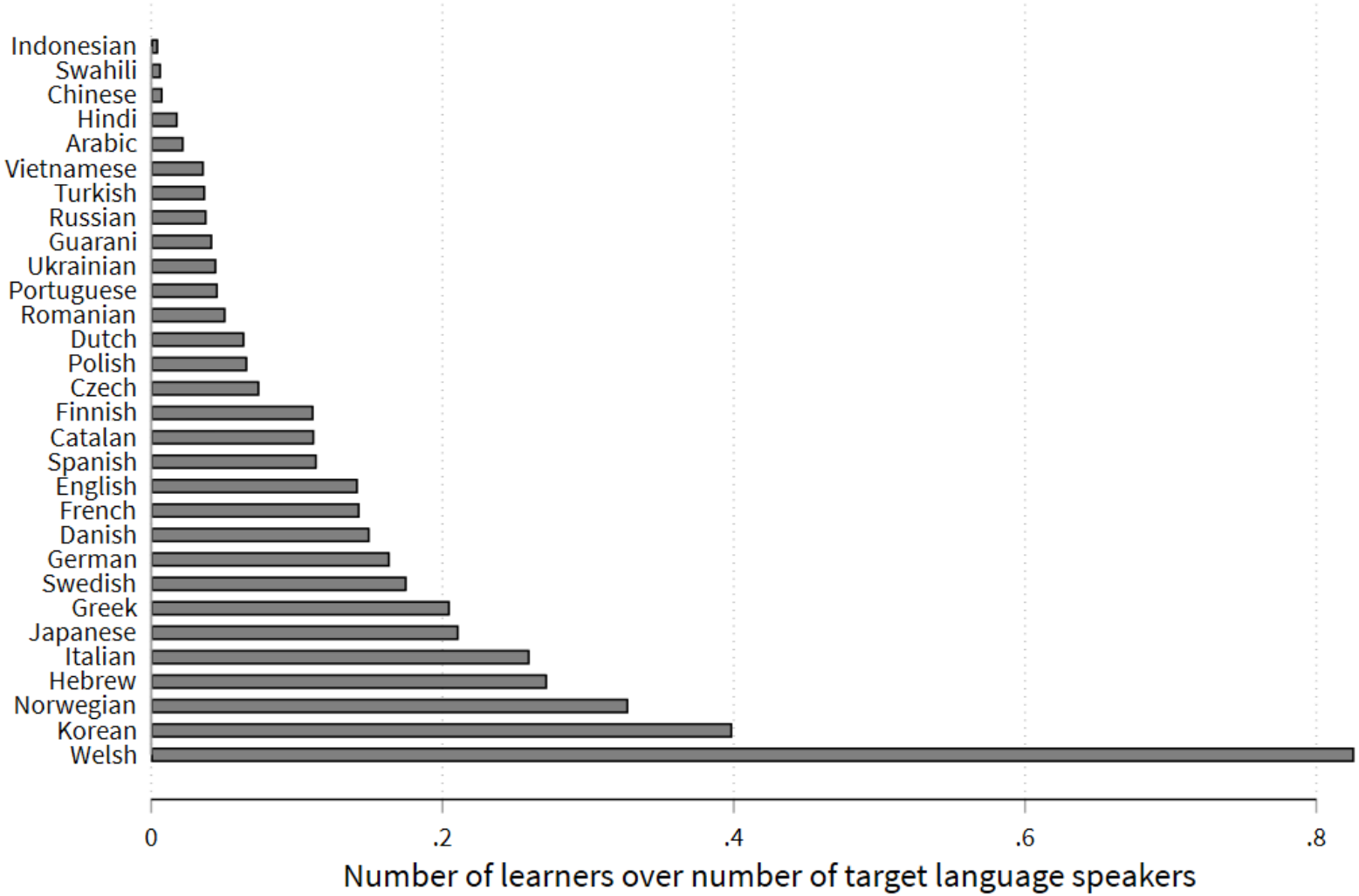
Estimated by OLS, Standard errors clustered at the country of origin.

Appendix

Number of learners by source language



Number of learners by target language



A model of low-cost language learning

Modified Random Utility Model (McFadden '74) with investments in foreign language T for someone from o, where α_{oS} speak S

- Period 1: Acquire language skill s_{oT} at convex cost $c_{oT}s_{oT}^2$, $c_{oT} = \kappa_{oT}/2(1 + \eta_l \alpha_{oS} Duolingo_{ST})$
- Period 2: Choose location. Utility in d: $U_{iod} = \ln w_{iod} = \mu_{od} + s_{oT} b_{odT} + \epsilon_{iod}$; EVT-1 distributed

Period 2 utility maximization *given* s_{oT} gives a multinomial logit form:

$$\mathbb{P}_{od} = \frac{e^{\mu_{od} + s_{oT} b_{odT}}}{\sum_{d'} e^{\mu_{d'} + s_{oT} b_{od'T}}}, \quad \ln\left(\frac{\mathbb{P}_{od}}{\mathbb{P}_{oo}}\right) = \mu_{od} - \mu_o + (b_{odT} - b_{ooT}) s_{oT}$$

A model of low-cost language learning

Modified Random Utility Model (McFadden '74) with investments in foreign language T for someone from o, where α_{oS} speak S

- Period 1: Acquire language skill s_{oT} at convex cost $c_{oT}s_{oT}^2$, $c_{oT} = \kappa_{oT}/2(1 + \eta_l \alpha_{oS} Duolingo_{ST})$
- Period 2: Choose location. Utility in d: $U_{iod} = \ln w_{iod} = \mu_{od} + s_{oT} b_{odT} + \epsilon_{iod}$; EVT-1 distributed

Period 2 utility maximization *given* s_{oT} gives a multinomial logit form:

$$\mathbb{P}_{od} = \frac{e^{\mu_{od} + s_{oT} b_{odT}}}{\sum_{d'} e^{\mu_{d'} + s_{oT} b_{od'T}}}, \quad \ln\left(\frac{\mathbb{P}_{od}}{\mathbb{P}_{oo}}\right) = \mu_{od} - \mu_o + (b_{odT} - b_{ooT}) s_{oT}$$

Optimal language skills in the low migration limit:

$$s_{oT}^* \approx \left(\mathbb{P}_{oo} b_{ooT} + \sum_{d \neq o} \mathbb{P}_{od}(0) b_{odT}\right) \frac{1 + \eta_{oT} \alpha_{oS} Duolingo_{ST}}{\kappa_{oT}}$$

$\rightarrow$ larger for languages with strong returns on domestic labor markets

From model to empirical strategy

$$\ln\left(\frac{\mathbb{P}_{od}}{\mathbb{P}_{oo}}\right) = \mu_{od} - \mu_o + (b_{odT} - b_{ooT})s^*_{oT}$$

- The sign depends on the relative strength of the domestic and foreign returns to language skills
- I proxy b's with the calculated probability that a Duolingo module enables communication between two random individuals, assuming language skills are randomly distributed among its population:

$$DL^{ST}_{od} = \mathbb{P}(comm_{od}|DL_{S\to T}, S) = \mathbb{P}(comm_{od}|S \wedge T) - \mathbb{P}(comm_{od}|S)$$

- Giving the following time-varying foreign and domestic exposures to Duolingo:

$$DL^{abroad}_{odt} = \max_{S,T} DL^{ST}_{od}\,\alpha_{oS}\,Duolingo_{STt}$$

$$DL^{domestic}_{oot} = \max_{S,T} DL^{ST}_{oo}\,\alpha_{oS}\,Duolingo_{STt}$$

- When multiple courses (S-T pairs) "bridge" two countries, take the course with the highest value

Limitations

- Share of local speakers may be a limited measure of returns (e.g. English)
- No heterogeneity in agents: prospective migrants have larger propensity to migrate $\mathbb{P}_{od}(0)$ and are thus more likely to take it up

Effects on migrant selection are ambiguous:

- Higher educated have propensity to migrate and are more likely to take-up learning
- Liquidity constraints: language learning becomes available for those for whom traditional language learning was too costly.

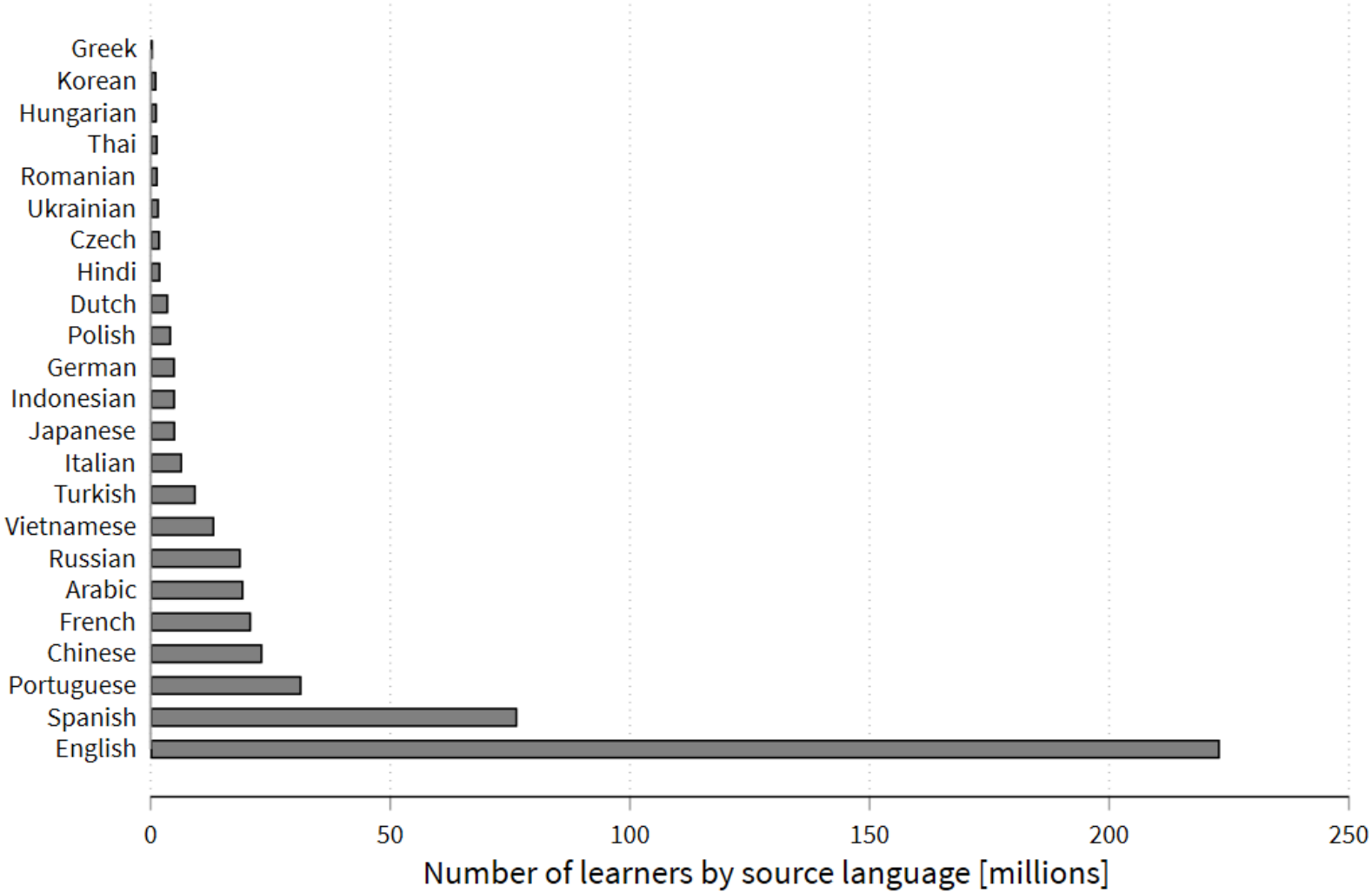
I do not model:

- Other learning motives (e.g. consumption motive learning for Korean)
- Post-migration learning (relevant for language skill acquisition in destination)

Determinants of rollout

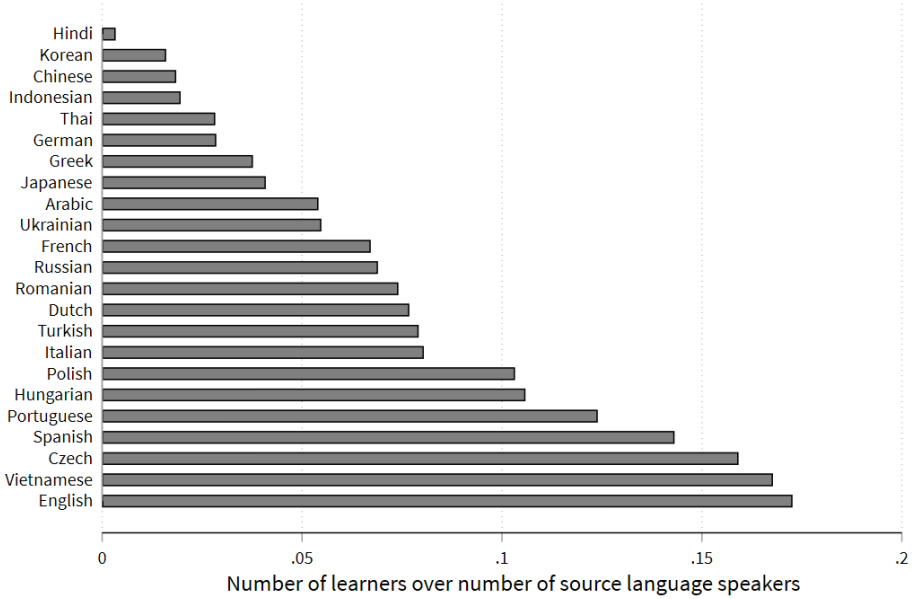
	$Duolingo_{23}^{ST}$		Year of rollout		$Duolingo_{od23}$		Year of large rollout	
Log source speakers	0.004**		-0.051					
	(0.002)		(0.265)					
Log target speakers	0.003**		-0.447**					
	(0.002)		(0.175)					
Log source speakers × Log target speakers		0.002**		0.460				
		(0.001)		(0.267)				
Sharing an official language					-0.198***	-0.205***	-1.393	0.751*
					(0.026)	(0.026)	(1.454)	(0.390)
Log distance					0.053***	-0.022***	-0.272	0.295***
					(0.009)	(0.006)	(0.221)	(0.071)
Log GDP pc PPP in origin					0.038***		1.380**	
					(0.009)		(0.531)	
Log GDP pc PPP in destination					0.026***		0.501**	
					(0.008)		(0.199)	
Log migrant stock + 1 (2005)					0.009***	-0.000	-0.104**	0.014
					(0.002)	(0.002)	(0.047)	(0.014)
Observations	13225	13225	84	52	22005	22005	8934	8912
Source and Target FE		✓		✓				
Origin and Destination FE						✓		✓

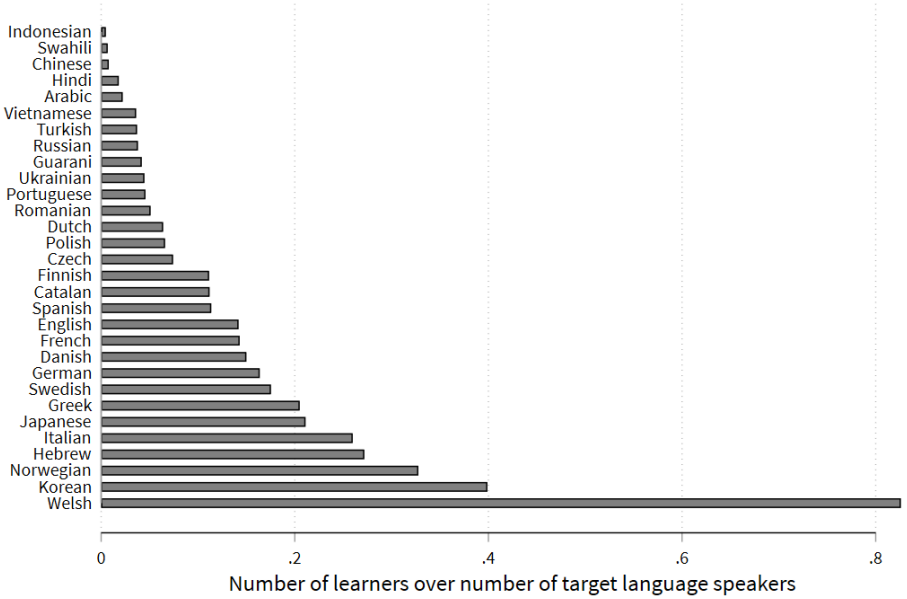
Number of learners by source- and target language



Number of learners by source- and target language

As a share of a language' total number of speakers





Determinants of uptake

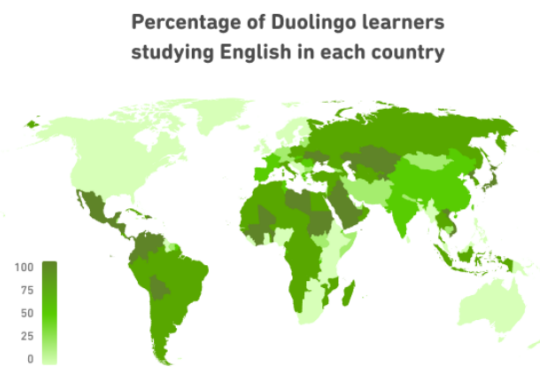
	(1) Learners	(2) Learners
Source language speakers	0.008***	0.004***
	(0.001)	(0.000)
Target language speakers	0.008***	0.004***
	(0.001)	(0.000)
Source language speakers × Target language speakers		0.000***
		(0.000)
Observations	84	84
Source and Target FE		✓

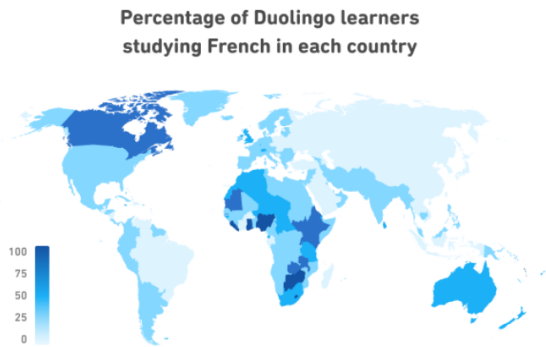
OLS regressions of the number of learners on Duolingo, as measured of the number of learners by language course, on the number of speakers of the source and the target language. Data on learners is obtained from the Duolingo platform in July 2024. Standard errors are clustered twoway on the source and destination language. The total number of learners is 478 million.

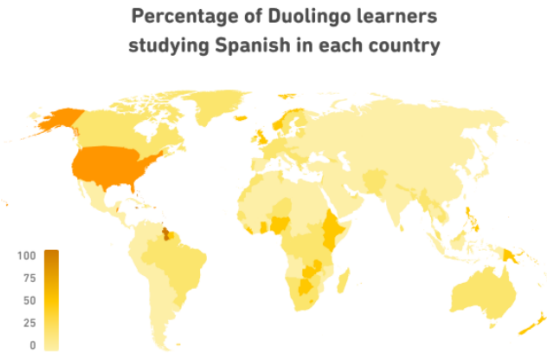
Duolingo: who learns what?

Source: *Duolingo* (2021)

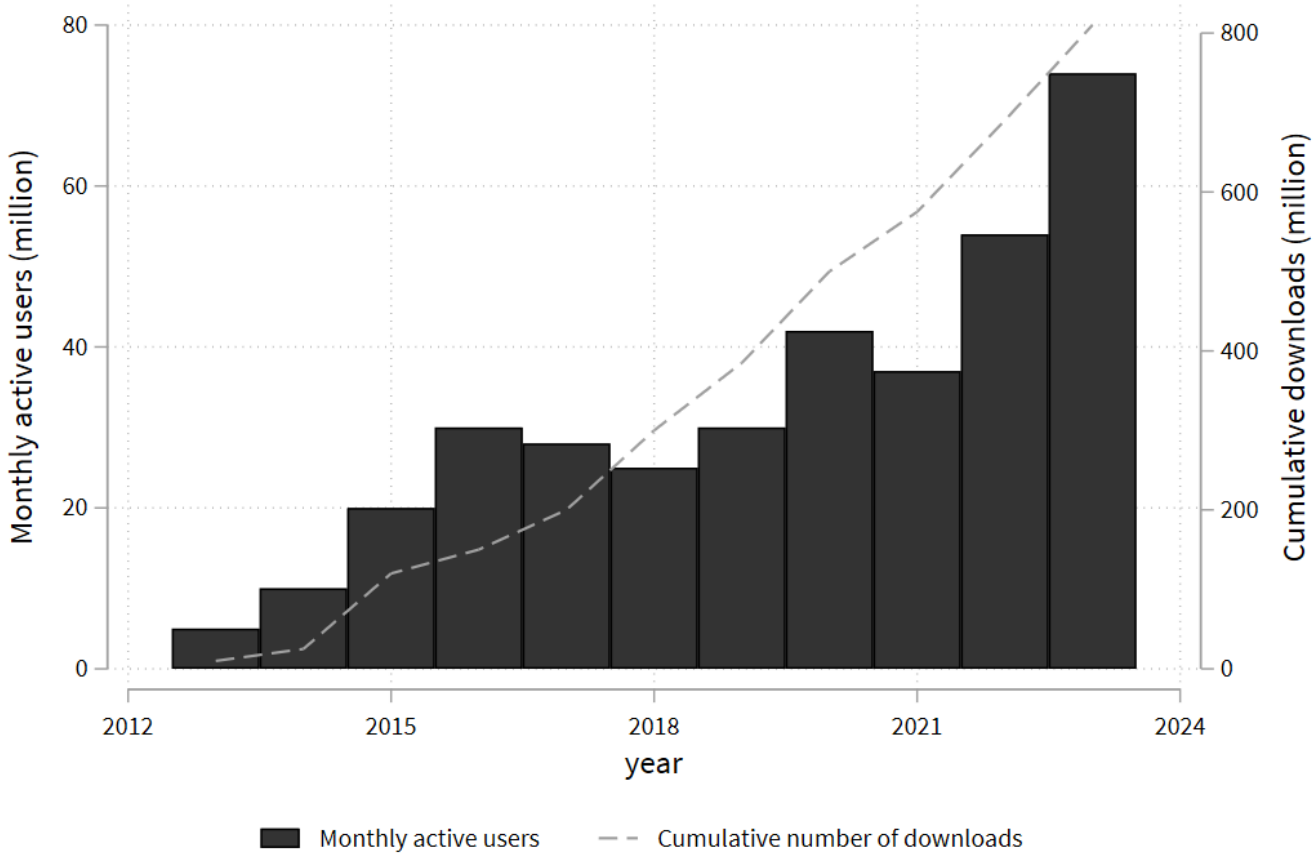
Percentage of learners learning English, Spanish, or French across the world in 2020







Growth in Monthly Active Users



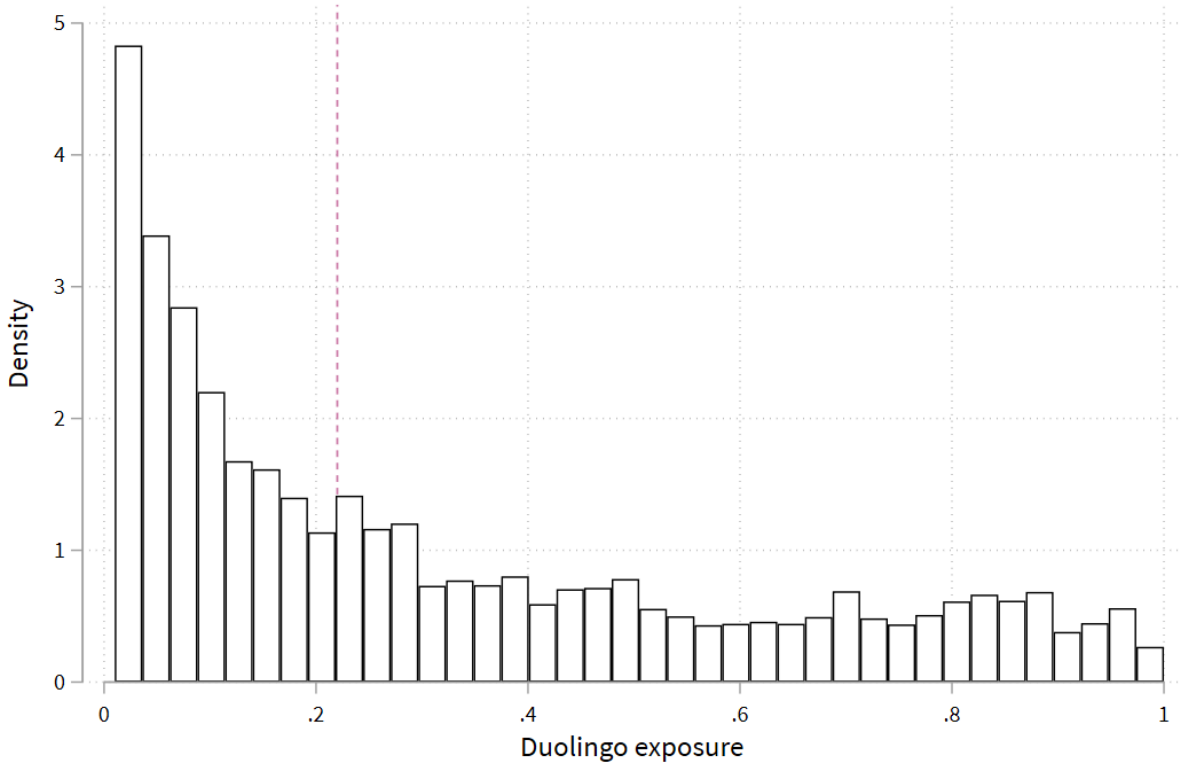
Numbers on the Monthly Active Users and cumulative downloads are obtained from `https://www.businessofapps.com/data/duolingo-statistics/`.

Traffic

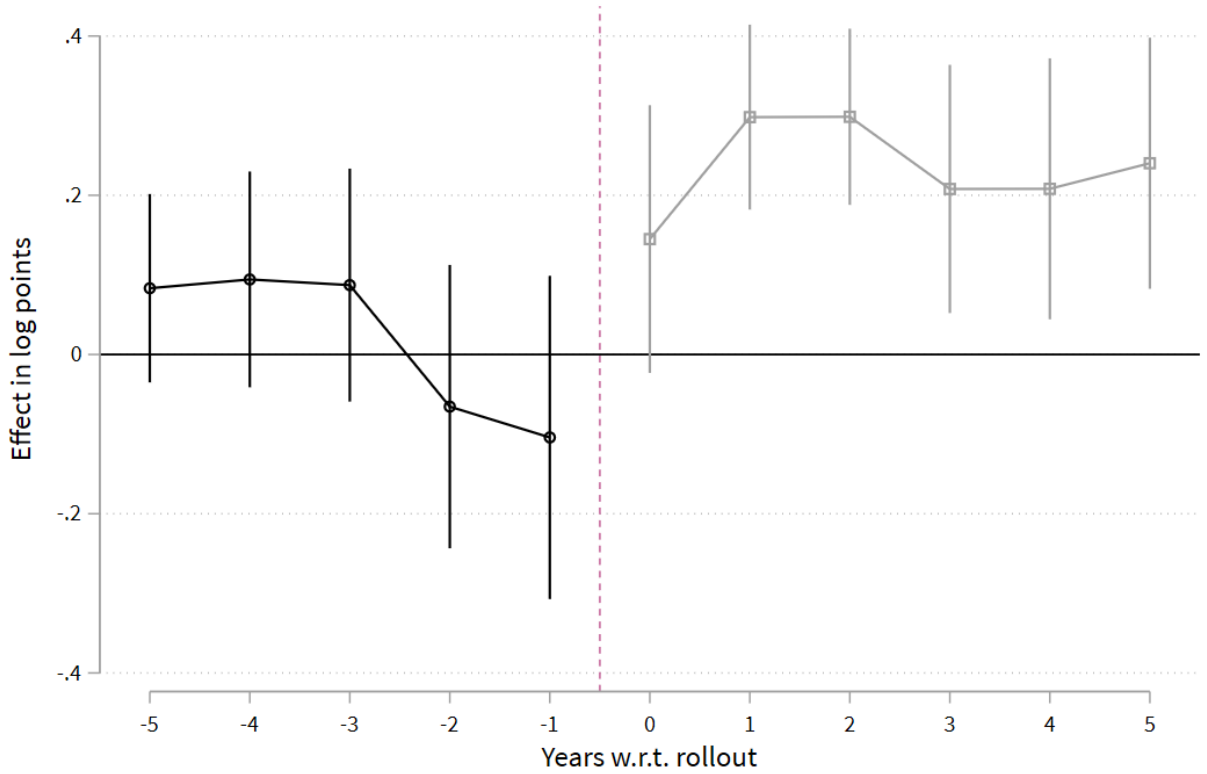
Global region	Share of traffic
North America	27%
South America	12%
Western Europe	11%
Eastern Europe	8%
Northern Europe	7%
East Asia	6%
Southern Europe	5%
South East Asia	5%
Central America	4%
Other	15%

Notes: Data has been obtained from Semrush (`https://de.semrush.com/website/duolingo.com/overview/`) in June 2024. Other includes Africa, Middle East and Turkey, and Oceania. Traffic from these regions is too low to analyze in isolation, but together accounts for about 15% of all traffic.

Exposure to *Duolingo*



Event study using Yotov-Nagengast estimator



Different ways of clustering standard errors

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
DL_{odt}^{abroad}	0.270***	0.413***	0.270***	0.413***	0.270***	0.413***	0.270***
	(0.066)	(0.055)	(0.073)	(0.087)	(0.077)	(0.030)	(0.033)
$DL_{odt}^{domestic}$	-0.422***		-0.422**		-0.422***		-0.422***
	(0.096)		(0.193)		(0.072)		(0.025)
Observations	114404	114404	114404	114404	114404	114404	114404
Lowest number of clusters	9855	9855	147	147	156	156	15
Unique dyads	9855	9855	9855	9855	9855	9855	9855
Origin-destination FE	✓	✓	✓	✓	✓	✓	✓
Origin-year FE		✓		✓		✓	
Destination-year FE	✓	✓	✓	✓	✓	✓	✓
Level of clustering	Pair		origin & destination		Linguistic pair		Linguistic

Omitting Covid-19 years

	(1)	(2)
DL_{odt}^{abroad}	0.273***	0.360***
	(0.068)	(0.065)
$DL_{odt}^{domestic}$	-0.378	
	(0.241)	
Observations	93819	93819
Unique origin countries	152	152
Unique destination countries	193	193
Unique dyads	9718	9718
Origin-destination FE	✓	✓
Origin-year FE		✓
Destination-year FE	✓	✓

Using official destination languages

	(1)	(2)
DL_{odt}^{abroad}	0.167***	0.198***
	(0.056)	(0.061)
$DL_{odt}^{domestic}$	-0.288	
	(0.327)	
Observations	123876	123876
Unique origin countries	153	153
Unique destination countries	194	194
Unique dyads	10694	10694
Origin-destination FE	✓	✓
Origin-year FE		
Destination-year FE	✓	✓

Monotonicity of effect

	(1)	(2)
$DL_{odt}^{domestic}$		
$0 < x \leq 0.2$	-0.000	
	(0.069)	
$0.2 < x \leq 1$	-0.199	
	(0.129)	
DL_{odt}^{abroad}		
$0 < x \leq 0.2$	-0.059	-0.001
	(0.054)	(0.054)
$0.2 < x \leq 0.4$	-0.033	0.152**
	(0.071)	(0.066)
$0.4 < x \leq 0.6$	0.093	0.222***
	(0.060)	(0.071)
$0.6 < x \leq 0.8$	0.119**	0.243***
	(0.061)	(0.072)
$0.8 < x \leq 1$	0.134*	0.321***
	(0.072)	(0.100)
Observations	124878	123484

Excluding high-income native English destinations

Excluding:	(1) AU	(2) CA	(3) US	(4) UK	(5) IE	(6) SAF	(7) all
DL_{odt}^{abroad}	0.377***	0.376***	0.384***	0.361***	0.378***	0.382***	0.320***
	(0.073)	(0.074)	(0.075)	(0.072)	(0.071)	(0.071)	(0.088)
Observations	96455	96449	96431	96442	96753	96979	89379
Origin-destination FE	✓	✓	✓	✓	✓	✓	✓
Destination-year FE	✓	✓	✓	✓	✓	✓	✓
Origin-year FE	✓	✓	✓	✓	✓	✓	✓
Origin-destination clustered SEs	✓	✓	✓	✓	✓	✓	✓

Alternative treatment definitions

	(1) M_{odt}	(2) M_{odt}	(3) M_{odt}
$DL_{\text{omit d-t pair with most speakers per lang.}}$	0.180***		
	(0.054)		
$DL_{\text{omit o-s pair with most speakers per lang.}}$		0.301***	
		(0.058)	
$DL_{\text{omit d-t and o-s pairs with most speakers per lang.}}$			0.254***
			(0.066)
Observations	100507	100507	100507
Origin-destination FE	✓	✓	✓
Destination-year FE	✓	✓	✓
Origin-year FE	✓	✓	✓
Origin-destination clustered SEs	✓	✓	✓

Migration policy

	(1)	(2)	(3)	(4)
DL_{odt}^{abroad}	0.400***	0.267***	0.315***	0.281**
	(0.083)	(0.085)	(0.104)	(0.127)
DL_{odt}^{abroad} × Permanent residence language requirements (0-2, MIPEX)			-0.053	-0.044
			(0.048)	(0.045)
DL_{odt}^{abroad} × English native high-income countries				0.051
				(0.113)
Observations	123484	41738	41738	41738

Estimated by PPML, Standard errors clustered two-way at the origin and destination level.

Language skills upon arrival

- 2021 EU LFS includes migrants' reported language skills *upon arrival*
- 2014 + 2021 EU LFS includes reason for migration
- $y_{odt} = DL^{T}_{odt-1} + \psi_{od} + \phi_{ot} + \theta_{dt} + \epsilon_{odt}$
- DL^{T}_{odt-1} is the Duolingo exposure to national language T
- Because EU LFS collects country of birth categories, we need to aggregate.

	At least beginner	At least intermediate	At least advanced	At least mother tongue
	(1)	(2)	(3)	(4)
Duolingo exposure	0.141***	0.125**	0.119***	0.010
	(0.050)	(0.054)	(0.033)	(0.021)
Observations	24115	24115	24115	28095
R^2	0.22	0.23	0.18	0.15
Average dependent variable	0.434	0.267	0.177	0.127

Language skills upon arrival

	Speaks EN	Speaks EN at least well	Speaks EN at least very well	Age at immigration	Female	At least 9 grade (at le 18)
	(1)	(2)	(3)	(4)	(5)	(6)
DL^{pre}_{oc}	0.009	0.020*	0.009	-0.342	-0.011	0.011*
	(0.010)	(0.012)	(0.011)	(0.264)	(0.008)	(0.006)
Observations	53671	53671	53671	67423	67423	61429
R^2	0.32	0.38	0.34	0.12	0.08	0.25
Average dependent variable	0.836	0.656	0.417	31.411	0.487	0.894
Fixed Effects	✓	✓	✓	✓	✓	✓
Controls for age, age2 and sex	✓	✓	✓			

Standard errors clustered at the region of birth level.

Reason of migration and tertiary education

	Employment	Job seeking	Family	Education	Refugees	Tertiary educated
	(1)	(2)	(3)	(4)	(5)	(6)
Duolingo exposure	-0.014	-0.089**	-0.011	0.057**	0.029	0.051*
	(0.025)	(0.041)	(0.053)	(0.022)	(0.032)	(0.028)
Observations	61045	61045	61045	61045	61045	225491
R^2	0.13	0.14	0.10	0.11	0.44	0.05
Average dependent variable	0.143	0.180	0.431	0.063	0.114	0.247

Does Duolingo Crowd out traditional language learning

Table 1: The Effect of *Duolingo* Courses on Institutional German Learning

	(1) Number of exams	(2) Number of course participants
Duolingo exposure	-0.019 (0.132)	-0.112* (0.058)
Observations	180	180

Columns 1 and 2 are estimated by PPML. Standard errors are clustered at the origin group level.

Worlwide interest in Duolingo and languages