

UC Riverside

UC Riverside Electronic Theses and Dissertations

Title

The Unequal Costs of the Dissemination of Knowledge: A Contemporary Case Study of English as the Scientific Lingua Franca

Permalink

<https://escholarship.org/uc/item/58k4t235>

Author

Kurtulmus, Taha Enes

Publication Date

2023

Copyright Information

This work is made available under the terms of a Creative Commons Attribution-NonCommercial License, available at <https://creativecommons.org/licenses/by-nc/4.0/>

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA
RIVERSIDE

The Unequal Costs of the Dissemination of Knowledge:
A Contemporary Case Study of English as the Scientific Lingua Franca

A Thesis submitted in partial satisfaction
of the requirements for the degree of

Master of Arts

in

Sociology

by

Taha Enes Kurtulmus

March 2024

Thesis Committee:
Dr. Steven G. Brint, Chairperson
Dr. Wei Zhao
Dr. Bekir S. Gür

Copyright by
Taha Enes Kurtulmus
2024

The Thesis of Taha Enes Kurtulmus is approved:

Committee Chairperson

University of California, Riverside

ABSTRACT OF THE THESIS

The Unequal Costs of the Dissemination of Knowledge:
A Contemporary Case Study of English as the Scientific Lingua Franca

by

Taha Enes Kurtulmus

Master of Arts, Graduate Program in Sociology
University of California, Riverside, March 2024
Dr. Steven G. Brint, Chairperson

Is the amount of research papers published in English by non-native English-speaking researchers determined by social conditions? To answer this, I use multiple regression analyses to study a random sample of academic researchers in Turkey (N = 3293). I show that it is determined by social conditions and that intra-national inequalities constitute the scientific lingua franca, i.e., English, as an additional dimension of the social hierarchy in the production of knowledge. The results suggest implementing state policies to mitigate the between-country inequality in engaging with international science and within-country inequality, which can potentially result from the former. The manuscript also demonstrates the necessity to treat academic English sociologically and to avoid the reductionism associated with applied linguistics.

Table of Contents

LIST OF TABLES	vi
INTRODUCTION	1
BACKGROUND	2
Literature review	4
METHODS	9
Data Collection and Sampling	9
Multiple Regression	11
Variables	11
Hypotheses	21
RESULTS	27
Multicollinearity	27
Regressions	28
DISCUSSION AND CONCLUSION	37
Hypotheses	37
Directions for Future Research	42

LIST OF TABLES

Table 1. Partial Correlations with High-Technology Exports in Current USD	4
Table 2. Regressions of the Approximate Percentage of Papers Published in English	30

INTRODUCTION

Science and technology are integral to economic development, national security, and public policy, and higher education institutions play a key role in the dissemination of scientific and technical knowledge for these purposes. English plays an important role in the production of this knowledge by the communicative efficiency it provides through access to an amount of knowledge not present in other languages, a terminology reflecting the latest research that is otherwise difficult to translate due to its sheer size and changing parts, and standardized linguistic practices that tend to clarify scientific communication. However, researchers are also typically bound to engage with academic English to avert scholarly isolation and due to the institutionalization of academic English as a response to its communicative efficiency. Yet to achieve this ‘communicative efficiency,’ researchers from non-Anglophone countries have to bear unequal costs in learning how to successfully communicate with it. These apply to disciplinary fields that are not tied to technological and economic outcomes as well, indicating the sheer importance of English as the scientific lingua franca.

While these unequal costs are clearer at the country-level, the intra-national inequalities among higher education researchers and teachers are understudied. This thesis is the first large-scale quantitative study that analyzes the impact of social conditions on academic English outcomes. The research questions I seek to answer are below, the first formulated more broadly and the second narrowed to partially operationalize the former:

1. What are the intra-national determinants of the differential engagement of non-Anglophone countries with scholarship internationally?
2. What are the social conditions determining the approximate percentage of papers published in English by non-native English-speaking researchers?

I begin by delineating the linguistic context in Turkey, which I take as my case, and then describe the state of research in applied linguistics, where the topic has been mostly studied. I introduce elements from Pierre Bourdieu as it relates to language and utilize it to develop causal hypotheses. I use multiple regression analyses to show the direct and implied indirect effects that social conditions (as well as organizational and disciplinary variables) have on English publications. After discussing my findings, I end with directions for future research.

BACKGROUND

The rise of English as the scientific lingua franca owes its existence to its uninterrupted status as an international language through British political imperialism and American economic supremacy (David 2003:7-10) and the decimation of much of Europe's scientific infrastructure during the Second World War (Kaplan 2011:9).¹

In historical Turkey, French replaced Italian as the primary diplomatic language in the early nineteenth century and the Ottoman Empire increasingly came under the Francophone sphere of influence during the Tanzimat Era that began with the Gülhane

¹¹ Vannevar Bush in his famous report to the US President: "We can no longer count on ravaged Europe as a source of fundamental knowledge" (Bush 1945:17).

Edict of 1839 (Davison 1994). French became firmly established particularly in modern schools (Kuşçu and Yağlı 2022; Özkan 2010), to be replaced by English only in the 1950s through emerging US-Turkish relations amid the reception of international aid (Selvi 2011:188). Despite its received importance in the country, Turkey ranks 66 out of the 113 countries measured in the EF English Proficiency Index for 2023, placed among the “low proficiency countries” (Education First 2023:6). Koru and Akesson's (2011:3) analysis that Turkey is underperforming based on economic indicators still seem accurate despite being published more than a decade ago.

For higher education organizations and systems, English is significant for its influence in international rankings—partially through its centrality to citation indexes and the necessity to offer programs in English to attract talented international students (Salmi 2009:61-63).² However, the impact of academic English likely extends beyond the educational realm and might have direct effects on economic and political economic outcomes.³ Considering the close relationship between basic and applied research and therefore economic outcomes,⁴ English (or more specifically, the academic English level of a country’s scientific and technological workforce) likely moderates the effect of scientific research and development on economic outcomes, since it provides linguistic access to a disproportionate amount of published research globally (Montgomery 2013)⁵

² For a number of university case studies in countries such as China, Singapore, Malaysia, Nigeria, and Chile, see Altbach and Salmi (2011).

³ While cultural outcomes and English might not have the same degree of mutually reinforcement between English and economic outcomes, English nevertheless has important effects on cultural power relations (at the individual and national levels), which is why I do not limit my disciplinary focus to STEM fields.

⁴ For the utility of maintaining the basic-applied distinction, see Roll-Hansen (2017).

⁵ Among five major languages, about 85% of research was indexed in English by the 1990s and more by the 2000s (Montgomery 2013:90). While the numerical bias contributed by the condition of being indexed

and ready set of terminologies not yet existing in other languages. In Table 1, a simple partial correlation shows that national scores of an aggregate H-index (Scimago Lab 2022) are significantly correlated with high technology exports in current USD.

Table 1.
Partial Correlations with High-Technology Exports in Current USD

Variable	Correlation	p-value	Degrees of Freedom	Observations
National H-index	0.5	0.03	17	23

Notes: The control variables are: (1) Gross capital formation in current USD; (2) GDP in current USD; (3) Educational attainment of population % (doctoral or equivalent); and (4) Research and development expenditure in current USD. "National H-index" is the Scimago Journal Rank aggregated at the country level for between 1996-2022. Control variables are from World Bank data from 2020. The chronological inconsistency is due to lack of data. The variable is still significant at $p < 0.05$ after controlling for population size.

While we do not possess sufficient aggregate data to show such a moderation (and even less at the level of academic and other researchers), the English for Academic Purposes (EAP) literature and others related to English as a lingua franca (ELF) have developed a body of work accounting for the unequal costs borne by non-native English-speaking researchers.

Literature review

English for Academic Purposes. EAP partly emerged from linguists' needs to descriptively study the different uses of contemporary English and increasingly disassociated itself from formal linguistics in the 1970s as it became increasingly

is clear, it should also be considered that access to indexed research can be easier compared to non-indexed research.

multidisciplinary through converging influences of oral discourse, discourse analysis, rhetoric, the emergence of the notion of the textual ‘genre’ in 1980s, and the rising social constructionism in the sociology of science and knowledge at the turn of 1980s (Swales 2001:47-48). Genre analysis, especially as developed by Swales (1990), was very influential in directing the attention of EAP and applied linguists to studying the textual genre of the ‘research paper,’ as well as issues of scientific language in general. A closely related and growing concern of the literature was the unintended malign effect of English. Influentially demonstrated by Swales’s (1997:374) famous “English as *Tyrannosaurus Rex*,” which highlighted the risk of English threatening the existence of special registers in other languages, this body of work grew to include the burdens it placed on non-native researchers during publication (e.g., Lillis and Curry 2010), on higher education teachers and students in university classrooms (e.g., Ammon 2001; Arkın and Osam 2015), and the unequal citations patterns as revealed by corpus analyses (e.g., Breeze 2015; Dontcheva-Navratilova 2015; Hewings, Lillis, and Vladimirova 2010).

While much of this research drew much from sociology and effectively carried out social research in the linguistic area (excepting for corpus analyses), it mostly does not treat language as a sociological object connected to and conditioned by other components in the social structure. Giving an adequate account of discourse and language cannot be achieved without also accounting for the conditions that produce the groups where they take place (Bourdieu 1977:650). In the few cases where researchers approximate such a conception of language, sociological analyses are either too limited, omitted, or lack methodological rigor.

For example, in one of the very few quantitative studies on non-native researchers' experiences with academic English across multiple disciplines, Medgyes and Kaplan (1992:72) only discuss 'age' as a social condition in Hungarian academia. Here, they do not discuss gender due to very low number female respondents, who were underrepresented in Hungarian academia at them time. In the follow-up study ten years later, age is still the only social condition discussed and data on gender is not even reported (Medgyes and László 2011:269, 282). Another study about Russia, exploratory in nature, also only reports on age (Kryuchkova 2001:415). The majority of quantitative studies are corpus analyses, where the integration of most important social conditions may be difficult or even impossible. However, such studies also abstain from including organizational variables derived from employer universities (relatively easy to collect from the metadata of the publications), tend to limit themselves to a small number of disciplines/departments, or aggregate disciplines into larger groups, effectively precluding the analysis of discipline-specific patterns (e.g., Breeze 2015; Oliver 2015). Others also do not implement significance tests (Bocanegra-Valle 2015; Hewings et al. 2010) to test their conclusions.

The disregard of social conditions is also present in qualitative research (e.g., Gnutzmann, Jakisch, and Rabe 2015; Kuteeva 2015), where it would be easier to gather data on participants' social background. McGrath (2014) and Schluer (2015) at least include academic seniority in their analyses, but they also omit pertinent social conditions (e.g., native language and gender) in order to maintain the anonymity of their respondents. Others in EAP have taken some interest in the interplay between race,

gender, and status (e.g., Starfield 2001), but these veer toward teaching (more specifically, English-medium instruction) rather than the role of English in research.

A consequence of applied linguists' focus on ELF as a linguistic object is the tacit elaboration of a principle of division into classes that separates one part of the world into the dominant and another into the dominated, usually through dichotomies such as "inner vs. outer/expanding circles, core vs. periphery scholars, native English speakers vs. non-native English speakers, Anglophone vs. non-Anglophone" (Plo Alastrué 2015:4). Many, if not most, EAP scholars and other applied linguists in the field work with these divisions in order to highlight and deconstruct the arbitrary power differentials between these dichotomic categories.⁶ In doing so, they reduce the problem of unequal burdens to a linguistic problem with two categories, effectively muting the differences and hierarchies within these categories. This is how one (larger) group of scholars are able to study a group of researchers or research output and come to the conclusion that non-native researchers are disadvantaged (most of the research cited above follow this line), while another (smaller) group can study other cases and come to the conclusion that English does not pose an additional challenge (Kuteeva and McGrath 2014).

Overcoming this tendency is important for two reasons, one at the individual and another at the country-level.

⁶ "[T]he different struggles over classifications, struggles over the monopoly of the power to make people see and believe, to get them to know and recognize, to impose the legitimate definition of the divisions of the social world and, thereby, to *make and unmake groups*." (Bourdieu 1991:221)

At the individual level, academic English is a specific type of linguistic capital for academic researchers and functions as an additional axis of hierarchization, insofar as there is linguistic profit to be made from it, which take the form of more citations (Hewings et al. 2010:106-107), fulfilling promotion requirements (Lillis and Curry 2010:55-56), better reception of research grant applications, and simply being able to avoid translation costs. Bourdieu's (1977, 1991) field theory and writings on language provide useful tools to analyze the social conditions of academic English at the individual level, which I will selectively incorporate into this study.

Bourdieuian field theory. Bourdieu (1991:654) argues that linguistic profit depends on two elements: (1) the unification of the linguistic market (effectively the mass recognition of the more valuable mode of language) and (2) the chances of producing legitimate language, which in turn depend on having linguistic capital and access to sites of expression. While the unification of the international market for academic English has mostly occurred since the Second World War, its degree depends on country and discipline. More immediate to researchers is the differential chances of producing academic English, which could function as a *de facto* censorship due to a smaller readership or decreased likelihood of international publishing in their native languages.⁷

Linguistic capital is a sub-type of the more general cultural capital, along with which it depends on an individual's place in the distribution of social resources. This

⁷ "An adequate science of discourse must establish the laws which determine who (*de facto* and *de jure*) may speak, to whom, and how . . . Among the most radical, surest, and best hidden censorships are those which exclude certain individuals from communication." (Bourdieu 1991:648)

suggests that we must consider direct educational capital (that is, of the researcher); educational capital transmitted by parents; characteristics of higher education organizations, since the struggle over the definition of legitimate culture is institutionalized in these (Bourdieu 1984:93); and ethnicity and gender-sex, due to the distribution of social resources between ethnicities and sexes (Bourdieu 1984:102) and the division of labor between the sexes (Bourdieu 1984:109), among other elements.

At the country-level, maintaining intra-national differences also sheds light on decreased engagement between scholars who primarily publish in different languages (i.e., English vis-à-vis the national lingua franca) or, at the very least, less engagement with research in the unpreferred language, which might indicate the potential formation of intra-national linguistic networks that interact with each other like separate countries. While this is not the concern of this study, it is worth mentioning as a closely related topic that has gone unstudied in the literature.

METHODS

I will use quantitative data as a way to predict the approximate percentage of papers published in English, which I conceive as a proxy for a given academic's position in the linguistic dimension of the social hierarchy of academics.

Data Collection and Sampling

The target population is the teachers and researchers in Turkish higher education. For data collection, I created a population frame from YÖK Akademik, a public repository of all currently employed higher education teachers and researchers at Turkish

universities, which is maintained by the Council of Higher Education (YÖK) of Turkey. I eliminated all individuals who were listed as *araştırma görevlisi* ('research employee'), a title meant to correspond to research assistants but is typically used as a legal category to employ non-academic staff workers. Using their publicly available contact information, I invited 50,500 individuals via email, who filled out self-administered online surveys between August 27 and November 6, 2023. The survey automatically appeared in Turkish to respondents due to their IP addresses, although they were able to switch the language to English. This enabled the participation of international academics as well. To reduce the completion time, certain variables that were publicly available (e.g., university, university type, department) were drawn from the same repository. The survey had a response rate of 11% and, among them, a completion rate of 63%. After data cleaning, the number of cases in the sample for this study has been reduced to 3293.

Although random sampling had been used for 50,000 of the contacts, an additional 500 were invited based on their employment status in six universities categorized as 'elite university' (also randomly sampled *within* that category). Accordingly, while academics from such institutions might be overrepresented by 1% in the sample, this does not have a negative bearing on the regression analyses as this group of academics is controlled as an independent variable.

Random sampling for this target population is also important for variance estimation because there is little known data about its features. Accordingly, researchers do not have many auxiliary variables to impose a balance on the sample (Tillé and Matei 2016).

Multiple Regression

I use multiple linear regressions to test multiple models. I hope to show the existence or non-existence of moderation through the iterations of most models. Two other models will be useful to see whether results change when a researcher might have different data considerations (whether self-reported data on language proficiency is inherently problematic) or may be constrained by data limitations (whether only publicly available data is present).

Variables

Dependent variable. I use the *approximate percentage of papers published in English* as my dependent variable, which was reported by participants from 0 to 100 in increments of 10. This only includes research papers. This variable may be considered as a partial measure of one's success in the linguistic dimension of the social hierarchy of academic research.

Independent variables. The independent variables are intended to reflect the social sources possessed by the respondent before earning the doctorate (or its equivalent), if the respondent earned a doctorate at all. For parsimony, I do not iterate in this study different models of only independent variables to see potential moderation between them, although this is a possibility for future research due to the chronological nature of the variables.

Gender-sex is the self-reported identification of the respondent's gender-sex, which was recoded as binary and refers to males as a dummy variable. I excluded the non-binary third value of "Other" due to only having 3 cases (0.1% of the sample).

Age refers to the respondent's age, calculated by subtracting their self-reported year of birth from 2023.

Ethnicity refers to the respondent's self-identification with an ethnic group, normally understood in terms of language and ancestry. Here, I use four dummy variables—Turk, Kurd or Zaza, Other domestic, and International. I have combined other domestic ethnic groups in the third dummy variable due to their small percentage. Here, 'domestic' refers to ethnic groups that have been present in Turkish territory since before the establishment of Turkey or have migrated en masse in the decades following its foundation (such as Assyrians, Lazi, Balkan ethnicities, and Circassians), who are most likely citizens of Turkey. While this precludes the analysis of these specific, it enables a comparison between the larger groups of the dummy variables. For the smaller models (1, 2, and 4), I use Ethnicity, international as the reference category. When multicollinearity becomes too large and affects significance, I use the same variable as the dummy variable and exclude all others, since together they form a meaningful category of domestic ethnicities opposite those who are international in origin.

The two variables of educational level refer to the highest level of education completed by the respondent's father and mother. From lowest to highest, the eight

ordinal levels are: ‘didn’t go to school,’ ‘primary school,’ ‘middle school,’ ‘high school,’ ‘associate degree (2 years),’ ‘bachelor’s (4 years),’ ‘master’s,’ and ‘doctorate.’

English preparatory school variables refer to whether the respondents attended an English preparatory school at the three possible stages of their education—(1) middle school, (2) high school, and (3) university. All three are binary variables. University-level prep schools are required for university programs in languages other than in Turkish, although those who demonstrate their proficiency in language exams can be excused from attending. High school-level prep schools are present in a small percentage of more selective high schools that usually require a much higher score in the university entrance exam. Middle school-level prep schools do not exist anymore, although there many academics who attended them before their closure. There is no reference category since prep schools at different levels are not mutually exclusive.

PhD origin variables refer to the regional origin of the respondent’s doctorate. This includes researchers who are enrolled in but have not yet completed a doctoral program. Due to the large number of countries or country-level regions (N=34), I use five dummy variables, the last of which is the reference category—(1) Turkey, (2) non-Anglophone European countries (ex. EE), (3) other countries abroad, (4) No PhD, and (5) Anglophone countries. Even though this last variable is not a geographical variable, it allows all cases to be exclusive and exhaustive, effectively functioning at the same level. The excluded reference category of Anglophone countries refers to Australia, Canada, England, Scotland, the USA, and Wales.

‘Non-Anglophone European countries (ex. EE)’ refer to European countries where a large portion of the population does not speak English as native speakers. This includes Germanic, Nordic, and French-speaking countries, as well as Spain, Italy, and Portugal. However, this variable also excludes Eastern European countries (represented by Slavic speaking countries in the sample), due to presumed differences in institutional performance metrics and in the prevalence of English-language programs in their higher education systems. ‘Other countries abroad’ refer to all other countries beyond these categories and Turkey. Despite the institutional diversity of the remaining countries abroad (which includes Turkic, Slavic, East Asian, Arabic-speaking, Hindu-Urdu speaking countries, as well as many others), academics from these regions share the feature of being international scholars, requiring the need to academically engage with English as the primary language of research. While those with doctorates from Turkic countries might pose a partial exception, they cannot by default rely on publishing in the Turkish of Turkey (*Türkiye Türkçesi*) in the setting of the regional origin of their doctorate (which use other Turkic languages), due to which they are treated similarly to Turkish nationals who study in other countries abroad.

Control variables. The control variables reflect different groups of circumstances that might moderate the effects of independent variables, such as organizational and disciplinary variables, or have an entirely unique influence on the dependent variable. Some of these are also relevant to engage with the literature. For the purposes of this study and to limit the number of hypotheses, I do not treat these as independent variables.

Organizational type refers to whether the respondent's employer university is public or has foundation status.⁸ This binary variable takes the value of 'foundation university' in the analyses. Foundation universities have more relative autonomy from the state through the board of trustees and do not primarily rely on financial support from the state. The employer university was drawn from the respondent's profile in YÖK Akademik and matched with its organizational type also listed in YÖK Akademik (Council of Higher Education 2023).

Elite status of university refers to whether the respondent's employer university is one of the nine universities I have designated as elite. These nine universities are: Koç University, İhsan Doğramacı Bilkent University, Sabancı University, TOBB University of Economics and Technology, Özyeğin University, Boğaziçi University, Middle East Technical University, Hacettepe University, and Istanbul Technical University. These universities appear at least three times on six different rankings that measure a variety of outcomes or similar outcomes with different methodologies. These are:

1. World University Rankings 2023 by Times Higher Education (2022),
2. University Ranking by Academic Performance by the URAP Research Laboratory (2022),
3. Universities with an average number of higher than two researchers supported in a TÜBİTAK-grant project (under the ARDEB program),⁹

⁸ 'Foundation universities' (*yakıf üniversiteleri*) are colloquially also called private universities (*özel üniversiteler*) in Turkey, although the latter does not have a legal existence as non-public universities can only be founded as non-profit entities through foundations.

⁹ TÜBİTAK is the primary governmental organization for funding scientific projects in Turkey, equivalent to the National Science Foundation in the US in function. This was manually calculated through the

4. Universities where more than 10% of their faculty are “superstar researchers” (Akçiğit and Özcan-Tok 2020:47),¹⁰
5. The fifteen universities with the highest score of research productivity adjusted by research quality (Akçiğit and Özcan-Tok 2020:38),
6. Universities with twenty or more international researchers (Yurdakul and Şahin Demir 2022:112)

The only exception that fulfilled these conditions but was left out is Gebze Technical University. Although it was one of only two universities that only had three appearances, it ranked lower in these and was excluded to make the group more homogenous.

Department groups variables are dummy variables that group the respondents according to their publicly listed departments on YÖK Akademik. This group of controls is necessary because (1) different programs and disciplines have large differences in how much skewed they are toward international publications and (2) some disciplines are content-wise inherently skewed towards international publications, such as programs about English language and literature.

information retrieved on the official search engine for research and research projects in Turkey, retrieved December 7, 2023 (https://search.trdizin.gov.tr/tr/proje/ara?q=&searchName=&order=title-ASC&facet-publication_year=2018&page=1&limit=100). There were a total of nine universities that passed the threshold. the average number was calculated to control for university size, as large universities naturally tend to surpass others in the absolute number of projects and researchers supported.

¹⁰ The authors of the report defined superstar researchers as those who published any amount of research in the most impactful 5% of journals in their respective fields. The journals were ranked by standardizing their average impact score through their standard deviations.

To make the 400+ unique department titles more manageable, I developed a classification scheme mostly based on the Classification of Instructional Programs (CIP) by the National Center for Education Statistics (2020), reducing the groups of departments to 44, along with a very small number of unclassifiable departments. Differences from the CIP mostly stem from (1) minor differences between the Turkish and US higher education conventions, (2) the need to further differentiate between language programs due to the focus of this study, and (3) the need to differentiate between departments that are disciplinarily and substantively similar but imply different places in the social hierarchy due to factors such as (3a) length of study, (3b) large differences in university entrance exam scores, and (3c) gaps in post-graduation earnings due to differences in professional-technician status in the field.¹¹

To make the departmental groupings more manageable, I include only nine that are scientifically, socially, and linguistically important— (1) engineering disciplines, (2) social science disciplines, (3) history or Turkological disciplines, (4) English-oriented programs, (5) physical or formal sciences, (6) medicine and related clinical sciences, (7) other engineering disciplines, (8) business, management, and marketing programs, and (9) biological and biomedical sciences. These dummy variables were either selected from among the 44 groups of departments or were made by combining them. These account

¹¹¹¹ An example is the “Medicine and related clinical sciences” category of the CIP. While my classification includes medical and pharmaceutical programs, the CIP also includes child development, speech therapy, nursing, and dentistry technician programs. In my classifications, such departments as classified as “Health supporting professions” to make the former category more homogenous. A similar occupational earning (3c) perspective is also shared by the International Standard Classification of Occupations (International Labour Organization 2008).

for 1,722 cases in the sample. I exclude all other departmental groupings from the analysis to function as an aggregated reference category. Even though this prevents a substantive comparison between the dummies and the reference category, the analysis still allows comparisons between these nine groups.

Engineering disciplines also includes computer science departments, which are separate in the CIP. Social science disciplines aggregate wide ranging departments such as economics and anthropology. I maintained this CIP convention as (1) the CIP does not have a humanities category and (2) separating some social science disciplines individually would have made them numerically too small. History and Turkological disciplines were combined for the purposes of this study. By Turkological disciplines, I refer to specialty fields that study the history, language, or culture of Turkish and Turkic cultures.¹² History departments were combined with these due to their overwhelming preference for studying Turkish history. English-oriented programs include English language teaching programs at the preparatory school level as well as language, literature, and translation programs that center English. Physical or formal sciences combine the ‘physical sciences’ and ‘mathematics and statistics’ categories of the CIP in order to have a larger sub-sample size. The two original groups were also relatively similar on several metrics (including linguistic patterns in publication), so as to reduce the arbitrariness of combining the two. Unlike the CIP, medicine and related clinical sciences here excludes health supporting professions that usually lead to occupations typically categorized as

¹² Examples include departments of ‘Turkish Language and Literature,’ ‘Contemporary Turkic Dialects and Literatures,’ and ‘Traditional Turkish Music.’

‘technicians’ rather than ‘professionals.’ Through the other engineering disciplines category, I distinguish between engineering professions associated with the Industrial Revolution 1.0 from more advanced industrial revolutions, which is captured in the first engineering disciplines variable.¹³ Unlike the CIP, business, management, and marketing programs excludes departments that typically lead to supporting services in these areas.¹⁴ Biological and biomedical sciences follows CIP conventions, with the exception that agricultural biotechnology and genetics have been placed under engineering disciplines.

Department language (Turkish to English) is an ordinal variable with three values that ranks a department’s official language of instruction from Turkish to mixed (English-Turkish) to English. For the sake of comparability between models, departments whose official language involves a language other than Turkish or English are excluded from this study.

Academic title refers to a respondent’s academic title at the time of data collection, which was also drawn from YÖK Akademik. From lowest to highest, these are: (1) ‘Guest researcher/instructor,’ (2) ‘assistant professor,’ (3) ‘associate professor,’ and (4) ‘professor.’ For the small number of cases where academics held a professorial title through fulfilling research conditions but were listed as instructors due to limited

¹³ Examples for other engineering departments include ‘Forestry Engineering,’ ‘Civil Engineering,’ ‘Machine Engineering,’ and ‘Textile Engineering.’

¹⁴ This dummy variable includes departments such as ‘Management’ and ‘International Trade’ but excludes those such as ‘Banking and Insurance,’ ‘Human Resources’ and ‘Logistics.’

quotas for appointment at their current departments, I coded these according to their professorial titles.¹⁵

I added three control variables that are partial measures of academic English competence. All three are Likert scales with five ranks: ‘never,’ ‘rarely,’ ‘sometimes,’ ‘often,’ and ‘always.’ Mental translation refers to how often the respondent translates her thoughts from Turkish to English before speaking English (as opposed to speaking without mental translations). Academic English competence in own field refers to the respondent’s self-reported competence of discussing academic topics in her own field of specialization. Use of Turkish-English dictionaries refers to the frequency of using Turkish-English bilingual dictionaries when reading or writing research in English. Participants answered this question in juxtaposition to the same question about monolingual English dictionaries. Preliminary analyses with other variables about linguistic competence (not produced in this study) showed that academics who had an easier time with academic English preferred monolingual English dictionaries more, and vice versa.

I added three control variables as partial measures for the respondent’s peer-review experiences with international journals in English.¹⁶ These have the same Likert scale structure as the previous three variables. Peer-reviewers correcting for syntax measures the frequency of receiving reviewer suggestions on grammar, syntax,

¹⁵ Such a researcher would be listed as “Instructor (Assistant Professor).”

¹⁶ For the purposes of non-Anglophone countries, an ‘international journal’ refers to any non-domestic journal that publishes generally or entirely in English. When publishing in international journals is a requirement for certain promotions, the status of ‘international journal’ might be conditioned on being listed in a scientific index.

misspellings, or punctuation. Peer-reviewers correcting for clarity measures the same for clarity, verbosity, and writing style. Peer-reviewers suggesting native-speaker assistance refers to the frequency with which reviewers suggest the respondent to seek the assistance of a native speaker.¹⁷ I expect that academic English competence (at least respondents' subjective self-perception of it) does not exactly correspond to publishing experiences in English due to the involvement of non-grammatical linguistic and extra-linguistic elements.

These two groups of variables—academic English competence and experience with peer-reviewer feedback—are not included to establish causality with the dependent variable. Since these are commonly used variables by applied linguists (except for ‘mental translation’) and are used usually to the exclusion of most or all social conditions, they will be useful to see whether they are sufficient explanatory variables.

Hypotheses

I hypothesize multiple relationships between the independent variables and the dependent variable.

There are two possible logics that might operate in the differential determination of published papers in English by the sexes. The first is present in the unequal distribution of resources. In Turkey, men have a higher rate of education at each level above primary school, have a higher employment rate than women for each level of

¹⁷ These three variables were taken from Bocanegra-Valle (2015:211, 223) with minor modifications for the first two.

education, are overrepresented in middle- and high-level managerial positions (79.3% vs. 20.7%), and are politically overrepresented in the Turkish parliament (82.7% vs. 17.3%) for the year 2021 (TurkStat 2023). Accordingly, male researchers should have more papers in English by this simple logic of social reproduction. Conversely, Bourdieu (1984:109) argues that women have precedence in matters of taste and culture according to the division of labor between the sexes. This observed general tendency that cuts across classes is even stronger among petty-bourgeois women, who have the highest amount of linguistic sensitivity and “hope to achieve social mobility by virtue of their capacities for symbolic production” (Bourdieu 1977:667). Accordingly, female researchers should have more papers in English by this logic of the division of labor between the sexes. These opposing logics are expressed below.

HYPOTHESIS 1.—*Being male increases the likelihood of a higher approximate percentage of published papers in English compared to women.*

HYPOTHESIS 2.—*Being female increases the likelihood of a higher approximate percentage of published papers in English compared to men.*

The EAP literature informs us that as English has further spread and replaced regional scientific lingua francas, a difference in the preference for publication language has emerged between older and younger scientists. Medgyes and Kaplan (1992:79) reported this for Members of the Hungarian Academy of Science, and a follow up study a decade later demonstrated the same pattern for self-reported academic competence (Medgyes and László 2011:269). Kryuchkova (2001:415) also demonstrated the same for

a chemical research institute of the Russian Academy of Sciences. This exemplifies what Bourdieu (1977:651) calls a ‘generalized linguistic devaluation,’ since the former regional scientific lingua francas of German and Russian have been increasingly devalued as the linguistic markets have unified with that of the US. Furthermore, academic promotions are tied to research output in Turkey, where publications in mostly-English indexes such as the SCI, SSCI, and AHCI are more heavily weighted.¹⁸ While the effects of such a generalized linguistic devaluation likely begun a long time ago, since English has been the primary foreign language in Turkey for about five decades (Selvi 2011:188), it is also likely that the relatively recent requirement of publishing in indexed journals have further lead to a devaluation of publications in Turkish. As in many other countries, academic promotions are tied to research output where publications in mostly-English indexes such as the SCI, SSCI, and AHCI are more heavily weighted.¹⁹ All of these predict that a lower age will be associated with a higher rate of English publications. A prediction in the same direction is implied by Bourdieu’s (1977:353-354) explanation that conservatism increases with age due to declining social positions (at least among some occupational classes). Whether construed as declining or simply not having any more space for promotions, it is possible that older and therefore academically established researchers feel less of a need to comply with burdensome

¹⁸ Universities have different equations as to how academic scores are calculated and weighted, although privileging these indexes is a common practice. For an up-to-date list of each university’s appointment criteria documents, where the calculations are presented, see Council of Higher Education (n.d.).

¹⁹ Universities have different equations as to how academic scores are calculated and weighted, although privileging these indexes is a common practice. For an up-to-date list of each university’s appointment criteria documents, where the calculations are presented, see Council of Higher Education (n.d.).

linguistic conventions and having to publish in indexed journals, which are often in English. This is hypothesized as below:

HYPOTHESIS 3.—*Lower age predicts a higher percentage of published papers in English.*

Ethnic origin may play a role on two different levels—one among domestic ethnicities, and a second between them and international scholars. Among domestic ethnicities, Turks may be presupposed to possess more resources as the largest ethnic group and therefore have a higher English publication rate by the logic of social reproduction. This may have been further augmented by the undecidability over the ethnic component of Turkish citizenship (Yeğen 2004).

Besides ethnic origin, international scholars are also differentiated by being (1) geographically cut-off from their native network of scholars, which might lead them to collaborate with colleagues in English; (2) being employed by universities whose native language they likely do not know, by which the common language of research output is only in English; and (3) possibly being hired more competitively due to smaller employment quotas for international applicants, which might select more productive researchers. Accordingly, international scholars are most likely to have a higher rate of English publications than their domestic colleagues.

These are hypothesized below.

HYPOTHESIS 4.—*Ethnic Turks have a higher approximate percentage of published papers in English compared to those from domestic minorities.*

HYPOTHESIS 5.—*Researchers from international backgrounds (i.e., non-domestic ethnic groups) have a higher approximate percentage of published papers in English compared to those from domestic ethnicities.*

I expect that having gone to English preparatory schools at different levels predicts a higher English publication rate simply because the participants would have spent more time being exposed to it. However, it is also the case that mostly selective middle and high schools tend (or tended) to have prep schools. Since selectivity in schools is not controlled for, it is possible that these variables might be biased in that direction. This might partly apply to university-level prep schools as well. While these are more common, it is possible that families prefer programs that have English-medium instruction for higher expected returns to their children. I must also note that the English preparation that took place in middle school in the past was postponed to the high school level in the 1990s, after which it was entirely abolished in 2005 (Koru and Akesson 2011:3). Even though these have been reimplemented in different ways in the last 10 years, this might to affect parts of the sample differently. To partially control for this, models including prep school always controls for ‘age.’ The hypothesis is as follows:

HYPOTHESIS 6.—*Having gone to a English preparatory school predicts a higher approximate percentage of published papers in English.*

Based on the logic of social reproduction, I also posit that parents with higher levels of schooling will yield higher rates of publications in English.

HYPOTHESIS 7.—*Higher parental education predicts a higher approximate percentage of published papers in English.*

Finally, I posit that the linguo-regional origin of one's doctorate influences English publication rates. Those who earned their doctorate in Anglophone countries should have a higher linguistic competence simply due to being exposed to English more. However, they might also have been exposed to research environments more conducive to publishing in indexed journals. Since higher impact journals tend to be located in Anglophone countries, it is plausible that the tacit (and explicit) skills involved in publishing are better disseminated throughout higher education organizations there. Together, this should lead to a higher rate of English publications. On the other hand, having earned a doctorate in Turkey should lead to a lower rate of English publications relative to a doctorate abroad because (1) Turkish nationals who earn a doctorate abroad experience a pressure to publish in English earlier due to graduate school requirements and (2) international scholars have to maintain a stronger portfolio of English publications to be employable in a foreign country. Even if the second reason is controlled through the ethnicity variables, the first should influence the relationship in this direction. Accordingly:

HYPOTHESIS 8.—*Having earned a doctorate in an Anglophone country predicts a higher approximate percentage of published papers in English compared to those from other regions.*

HYPOTHESIS 9.—*Having earned a doctorate from Turkey predicts a lower approximate percentage of published papers in English compared to those from abroad.*

RESULTS

Multicollinearity

No other set of variables or individual variable demonstrated multicollinearity other than the ethnic origin variables. *Ethnicity, Turk* has a variance inflation factor (VIF) of more than 10 consistently across models, while *Ethnicity, Kurd or Zaza* and *Ethnicity, other domestic* vary around and above 5 across models. The multicollinearity is likely due to the (1) very high percentage of ethnic Turks (88.8%), which results in very high negative intercorrelations between it and the other two dummy variables, and the (2) very low percentage of internationals (0.8%), since a low percentage of the frequency of the reference category leads to multicollinearity (Wissmann, Toutenburg, and Shalabh 2007).²⁰ Despite this, the ethnic variables mostly remained significant despite multicollinearity and changing the reference category eliminated a desirable balance of significance. Since multicollinearity presents the risk of undermining statistical significance (Allen 1997:176) but the ethnic variables remain significant in these models, I have retained the three dummy variables in the early models (1, 2, and 4). However, in the full causal and the largest models (10 and 11, respectively), multicollinearity becomes too large for any significant relationship, due to which I have excluded all three dummy

²⁰ Wissmann, Toutenburg, and Shalabh (2007:10-11) state that a reference category with 4% of the observations will lead to moderate multicollinearity. It is therefore understandable why *Ethnicity, Turk* has a VIF of more than 10.

variables and included the international ethnic dummy variable. In other words, I treat all domestic ethnic groups (presumably equivalent to ‘Turkish nationals’) as the reference group and *Ethnicity, international* as an independent variable so salvage a meaningful significant relationship.

Regressions

The results of the multiple regressions are in Table 2 below. All models were significant according to the *F*-test ($p < 0.05$).

Model 1 ($N = 2693$; $R^2 = .03$) displays the variables at play before the respondents could have attended prep school at any level. Gender-sex, Age, and ‘Ethnicity, Kurd or Zaza’ are insignificant, whereas the rest are significant. The difference between ‘Ethnicity, Turk’ and ‘Ethnicity, other domestic’ may not be comparable at a significant level due its small magnitude, yet we understand that domestic ethnicities considered together seem predict around a 16% decrease in English publications. Finally, parental educational levels are both significant, with mother’s education having a slightly stronger coefficient.

Model 2 ($N = 1823$; $R^2 = .03$) displays the same variables with the three prep school variables added. Effectively, this denotes the variables that are in effect before enrolling in the doctorate (if at all). Gender-sex and Age remain insignificant. ‘Ethnicity, Kurd or Zaza’ turns significant in this model, while the ethnic dummy variables together still have a negative coefficient but with larger coefficients. While being a Turk predicts higher a rate of publications than other domestic ethnicities, this is not the case for being a

Kurd or Zaza. For parental education, father's education ceases to be significant and mother's education is still significant albeit with a smaller coefficient. English preparatory school variables are significant for each of the three levels, with prep school at the middle school level having the largest effect on the dependent variable, followed by the university level.

Model 3 ($N = 3201$; $R^2 = .06$) compares the effects of the regional origin of one's doctorate. All four variables (Turkey, non-Anglophone European country, other country abroad, and not having a PhD) are significant and have negative coefficients, meaning that the reference category of Anglophone countries predicts the largest approximate percentage of English publications without controls. Furthermore, among the categories that denote having a PhD (or being enrolled in a program), Turkey predicts the lowest approximate percentage.

Model 4 ($N = 1807$; $R^2 = .07$) combines all of the independent variables (i.e., models 2 and 3). Gender-sex and Age are still insignificant, and father's education remains insignificant when computed along the prep school variables. Mother's education retains significance but experiences a small decrease in its effect. The prep school variables also retain their significance. However, unlike Model 3, PhD origin, non-Anglophone European country loses its significance along the other variable, while the other PhD-level variables are significant.

Table 2
Regressions of the Approximate Percentage of Papers Published in English

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9	Model 10	Model 11
Constant	57.48*** (8.02)	61.75*** (10.33)	81.12*** (2.23)	89.47*** (11.33)	46.71*** (1.92)	28.11*** (1.14)	35.42*** (4.03)	77.46*** (1.99)	41.46*** (3.11)	55.74*** (5.94)	61.26*** (7.78)
Gender-sex (male)	-0.59 (1.35)	-1.17 (1.61)		-2.21 (1.59)						-3.22* (1.33)	-3.38** (1.27)
Age	0.6 (0.08)	0.15 (0.1)		0.04 (0.1)						-0.66*** (0.1)	-0.62*** (0.1)
Ethnicity, Turk	-16.3* (6.65)	-23.79*** (8.52)		-21.89* (9.25)							
Ethnicity, Kurd or Zaza	-14.38 (7.32)	-22.14* (9.21)		-21.56* (9.88)							
Ethnicity, other domestic	-16.45* (7.38)	-24.51*** (9.29)		-23.1* (9.96)							
Ethnicity, international (REF)										21.66* (9.7)	16.93 (8.86)
Educational level, father	1.51** (0.46)	0.77 (0.55)		0.67 (0.55)						0.34 (0.45)	0.67 (0.44)
Educational level, mother	1.96*** (0.53)	1.56* (0.61)		1.28* (0.61)						0.68 (0.5)	-0.16 (0.48)
English preparatory school, middle school		9.76*** (2.34)		8.92*** (2.3)						5.65** (1.9)	2.05 (1.78)
English preparatory school, high school		4.71* (2.13)		5.14* (2.11)						5.11** (1.72)	1.78 (1.63)
English preparatory school, university		6.45*** (2.04)		5.81** (2.01)						6.18*** (1.66)	3.9* (1.57)
PhD origin, Turkey											-5.95** (2.24)
PhD origin, non-Anglophone European country (excludes EE)			28.51*** (2.32)	23.46*** (2.79)					15.37*** (2.02)	12.51*** (2.43)	-7.27 (4.93)
			-11.12* (5.36)	-8.14 (6.64)					-10.89* (4.36)	-9.92 (5.47)	

PhD origin, other country abroad	-18.89* (8.19)	-25.65* (11.76)				-5.7*** (6.6)	-12.48 (9.86)	-11.66 (8.99)
PhD, none	-	-				-	-	-7.93
	41.78*** (3.52)	32.12*** (4.94)				31.06*** (3.17)	22.51*** (4.42)	(6.32)
PhD origin, Anglophone (REF)			5.12*** (1.53)					
Organizational type (foundation)					23.65*** (2.36)			
Elite status of university								
Department: engineering disciplines					35.4*** (2.32)	31.36*** (2.32)	29.35*** (2.77)	21.91*** (2.49)
Department: social science disciplines					-2 (1.96)	-4.79* (1.95)	-6.34** (2.41)	-8.62*** (2.3)
Department: history or Turkological disciplines					-	-	-	-18.14*
					29.98*** (2.83)	30.85*** (2.8)	30.49*** (4.48)	(7.99)
Department: English-oriented programs					23.32*** (2.48)	31.95*** (2.73)	31.7*** (3.35)	17.76*** (3.31)
Department: physical or formal sciences					48.63*** (2.84)	46.07*** (2.79)	41.34*** (3.59)	32.28*** (3.23)
Department: medicine and related clinical sciences					29.1*** (1.56)	28.18*** (1.57)	23.38*** (2.1)	18.17*** (1.9)
Department: other engineering disciplines					27.66*** (2.22)	24.31*** (2.22)	23.81*** (2.79)	17.06*** (2.5)
Department: business, management, and marketing programs					24 (2.5)	2.04 (2.51)	2.67 (3.16)	-0.76 (3.04)
Department: biological and biomedical sciences					31.94*** (2.56)	30.05*** (2.55)	28.22*** (3.27)	21.88*** (2.94)
Departments: other (REF)								
Department language (Turkish to English)					11.42*** (0.76)	8.66*** (0.84)	7.48*** (1.04)	4.29*** (0.99)

Academic title (junior to senior)									1.71**	6.59***	1.91*
									(0.52)	(0.85)	(0.86)
Mental translation											0.39
											(0.69)
Academic English competence in own field											6.15***
Use of Turkish-English dictionaries											(0.8)
Peer-reviewers correcting for syntax											-2.05***
Peer-reviewers correcting for clarity											(0.74)
Peer-reviewers suggesting native-speaker assistance											0.19
											-2.19*
											(0.96)
											-3.78***
											(0.95)
											1.29*
											(0.62)

R-squared	0.03	0.03	0.06	0.07	0.03	0.36	0.2	0.03	0.39	0.4	0.39
Observations	2693	1823	3201	1807	3230	3225	2825	2478	3095	1744	1407

Notes: Standard errors are presented in parentheses. Reference categories are denoted with an "(REF)" at the end.
 * p < .05 ** p < .01 *** p < .001

Model 5 ($N = 3230$; $R^2 = .03$) only includes the organizational control variables of Organizational type (i.e., ‘foundation’ as opposed to ‘public’ status) and Elite status of university. Both are significant and have positive coefficients. Perhaps unsurprising but important is the largeness the coefficient of Elite status: 23.65.

Model 6 ($N = 3225$; $R^2 = .36$) includes the departmental control variables, all but one of which denote programs and disciplinary groups, whereas the last denotes the official language of the program. All department groups are significant ($p < 0.001$) except for the Social science disciplines and Business, management, and marketing programs groups. The three department groups with the highest predictions of English publications are Physical or formal sciences (48.63), Engineering disciplines (35.4), and Biological and biomedical sciences (31.94). The three groups that follow are Medicine and related clinical sciences (29.1), Other engineering disciplines (27.66), and English-oriented programs (23.32). The group with significance that predicts the lowest rate of English publications is History or Turkological disciplines. The departmental variables explain much more variance than the previous models.

Model 7 ($N = 2825$; $R^2 = .2$) only includes three control variables that denote linguistic competence. Mental translation is the only insignificant variable, which is interesting as the variable had one of the strongest loadings on exploratory factor analyses intended to construct a latent variable of linguistic competence (not reproduced in this study). Academic English competence in own field is significant and has the sizeable coefficient of 10.34. Use of Turkish-English dictionaries is also significant and is associated with a 4.66 decrease in the approximate percentage of English publications.

However, it cannot be said that these cause approximate rates of publishing in English as they reflect the participants' current level of English and are concurrent with the publication percentages.

Model 8 ($N = 2478$; $R^2 = .03$) only includes three control variables that denote researchers' experience with peer-reviewer feedback. All three variables are significant and the first two—Peer-reviewers correcting for syntax and Peer-reviewers correcting for clarity—have negative coefficients as expected. Unexpectedly, Peer-reviewers suggesting native-speaker assistance has a positive coefficient, even if it has a lower absolute value than the first two variables. Correlations and partial correlations between these three variables (not reproduced here) show that all three are positive correlated as expected. However, as in the regression, Peer-reviewers suggesting native-speaker assistance is also negatively correlated with English publication rates even without controlling for the first two peer-reviewer variables (not reproduced here). This model also has a modest R^2 like the earlier models, indicating that self-reported peer-reviewer feedback alone does not explain much variance in the percentage of English publications.

Model 9 ($N = 3095$; $R^2 = 0.39$) combines all variables whose data can theoretically be collected through publicly available information. This includes regional origins of the doctorate, organizational variables, department groups, department language, and Academic title. Unlike Model 4, where regional origins of the doctorate variables were regressed along with the so-called pre-PhD variables, all four dummy variables of the regional origins are significant. Their coefficients are similarly negative; however, they have all decreased in absolute value when controlling for the other

variables in the model. Unlike Model 5, where only organizational variables were used, Organizational type is insignificant. Elite status of university is still significant, albeit, unsurprisingly, its former large coefficient of 23.65 has decreased to 7.23. Among the department groups, Business, management, and marketing programs remains insignificant, whereas Social science disciplines has become significant and has a negative coefficient. There is little change in the order of the size of the coefficients. Otherwise, Social science disciplines predict the second least amount of rates of English publication, after History or Turkological disciplines. Department language is still significant. Academic title, included for the first time in this model, is significant and positive, indicating that higher academic titles predict a modest increase (1.71) in English publications rates.

Model 10 ($N = 1744$; $R^2 = 0.4$) is the full causal model. It includes all variables groups except for those measuring linguistic competence and experience with peer-reviewer feedback, since these have no causal relation to the dependent variable. Here, Gender-sex and Age become significant for the first time. Being male predicts a decrease in English publication rates by 3.22 and age by 0.66. Here, due to multicollinearity, I have excluded the ethnicity dummy variables to function as a reference group, representing all domestic ethnicities. 'Ethnicity, international' has a coefficient of 21.66. Unlike previous models, both parental educational variables are insignificant after introducing controls. All three prep school variables remain significant even after the controls, although the larger coefficient of the middle school level from the previous models is now at a similar level with the other levels. Among the regional origin of the

doctorate variables, only the Turkey and PhD, none variables are significant. Although PhD origin, non-Anglophone European country became insignificant after the introduction of the so-called pre-PhD variables, PhD origin, other country abroad has become significant for the first time. Like the previous model, Organizational type is insignificant and Elite status significant. The department groups also follow the same pattern of significance as the previous model, where Business, management, and marketing programs is the only insignificant dummy variable. Academic title is also significant. This model is not as parsimonious as Model 9.

Model 11 ($N = 1407$; $R^2 = .39$) is the most highly saturated model and incorporates the linguistic competence and peer-reviewer feedback variables into the full causal model in Model 10. It functions to show whether only looking at these additional variables, which are the primary empirical foci of applied linguists, are sufficient in explaining the approximate percentage of papers published in English.²¹ Gender-sex, Age, PhD origin, Turkey, Elite status of university, all eight previously significant department group variables, Department language, and Academic Title remain significant from the full causal model. However, 'Ethnicity, international,' prep schools at the middle school and high school levels, and not having a PhD have become insignificant after the introduction of the linguistic competence and peer-reviewer feedback variables. These variables, in turn, have retained their significance, except for Peer-reviewers

²¹ I qualify this criticism to exclude those who exclusively conduct corpus analyses. Other applied linguists might also collect data on departmental affiliation yet these do not tend to be sufficiently comparative. However, applied linguists may plausibly object to the implication that approximate percentages in English publications do not properly reflect the challenges experienced by non-native English-speaking researchers.

correcting for syntax. Peer-reviewers suggesting native-speaker assistance, whose coefficient was unexpectedly positive in Model 8, still remains positive. This model is not as parsimonious as Model 10.

DISCUSSION AND CONCLUSION

Hypotheses

In discussing causality, I disregard Model 11 unless I also seek to establish a non-linguistic causal relationship, as I developed this model for the specific purpose to see whether social conditions were still significant after the inclusion of the linguistic competence and peer-reviewer feedback variables.

Hypotheses 1 and 2 posited opposite arguments about which gender-sex predicted a higher percentage of published papers in English. Model 10 confirms that being female is associated with an increase in the dependent variable (Hypothesis 2). I reject Hypothesis 1, whereas I fail to reject Hypothesis 2. Accordingly, it seems that the division of labor between the sexes is a stronger mechanism in this realm than a straightforward social reproduction through more resources. However, it must also be noted that significance was obtained only after the inclusion of the control variables (compare with models 1, 2, and 4).

Hypothesis 3 posited that a lower age is associated with a higher rate of papers in English. Model 10 confirms that as one gets older, there is a 0.66 predicted decrease in the approximate percentage of English papers. Accordingly, I reject the null hypothesis. The ‘generalized linguistic devaluation’ argument is also a geopolitical argument.

However, since the unification of the linguistic market has occurred several decades ago, it is possible that its effects have diminished or have been nullified. Regardless, it is not possible to distinguish between it and the ‘declining social positions’ mechanism with the models at hand.

For Hypothesis 4, I partially reject the null hypothesis, as ethnic Turks have a lower predicted rate of English publications than ethnic Kurds or Zazas but a higher one than other domestic ethnicities. Furthermore, the very small differences in the coefficients and the aggregate categories (the effect of which is suggested by the relatively large standard errors) indicate the need for more refined models. However, there is a possible explanation for the ethnic Kurd or Zaza category having the highest predicted value among all domestic ethnicities. There are no recent official statistics on ethnicity, but a survey from 2013 using multiple sampling methods showed the distribution of “mother tongues” in Turkey as the following: 80.4% Turkish, 16.4% Kurdish, and 3.2% other languages (Hacettepe University Institute of Population Studies 2015:62). The ethnic distribution of the corresponding groups in the full sample are as the following (excluding internationals to account only for domestic groups): 88.8% Turks, 5.7% Kurds and Zazas, and 4.6% other ethnicities. Accordingly, it is theoretically possible that the underrepresentation of this group (the specificity of which is made difficult by its aggregate grouping) might indicate a higher selectivity in those who joined the academic workforce. This selectivity might be due to other social conditions not explored in this study. It is not associated with a higher level of parental education, as

bivariate correlations (not reproduced here) show that they have the lowest averages among the domestic groups.

For Hypothesis 5, the null hypothesis is rejected through models 1, 2, 4, and 10. Notwithstanding a single insignificant dummy variable in Model 1, the international backgrounds of researchers consistently predict a higher rate of publication in English, which is 21.66 in Model 10. This is likely due to both (1) a network effect where international scholars are constrained to publish in English and (2) a higher selectivity among internationally mobile researchers. This finding justifies the efforts of higher education administrators who pursue international scholars to increase their university's standing. Model 11 might seem to suggest that this is not the case since 'Ethnicity, international' loses significance after the inclusion of linguistic competence variables. However, the insignificant p-value of 0.056 in Model 11 decreases to 0.045 and becomes significant after the removal of the Use of Turkish-English dictionaries variable. This variable is not pertinent to international researchers but still reduces its p-value either through a slight decrease in the observations for 'Ethnicity, international' or another undetected effect.

For, Hypothesis 6, the null hypothesis is rejected through all models that incorporate the prep school variables.

For Hypothesis 7, the null hypothesis partially rejected. Parental educational levels mattered differently both between the parents and at different stages. The father's educational level losing significance from the shift from Model 1 to Model 2 suggests

that its impact may be moderated by participants' education before enrolling in a doctoral program. In other words, the father's educational level may have an indirect impact on the dependent variable through a participant's pre-doctoral education. A moderation is also implied by the model for the mother's educational level. However, it maintained a direct influence on the dependent variable even after including the regional origin of the doctorate variables (Model 4). This is important, considering that researchers have usually already begun publishing at this point. Comparatively, the mother's educational level also consistently had a larger coefficient than the father's across models. This is in line with cross-national data on the larger impact of mother's educational levels on student performance (Marks 2008:303).

For Hypothesis 8, the null hypothesis is partially rejected. Having earned a PhD in an Anglophone country consistently predicts a higher value in the dependent variable in each significant dummy variable in this group. Even though some of the dummy variables were insignificant in different models, these are mostly the variables about having earned a doctorate in other countries (i.e., not Turkey), whereas PhD origin, Turkey is always a negative predictor. Since international scholars are already controlled for through 'Ethnic, international,' this suggests that researchers from domestic ethnicities who study abroad do not significantly differ in English publications from those who studied in Anglophone countries. This might suggest that international mobility is inherently likely to increase publishing in English due to the same network effect, mentioned when discussing Hypothesis 6.

For Hypothesis 9, the null hypothesis is rejected through all models where the variables for the regional origin of the doctorate were included. For each significant dummy variable, having earned a doctorate in Turkey predicts a lower rate of publishing in English. This is even valid when controlling for linguistic competence in Model 11, suggesting that graduate-level Turkish higher education might be suffering from a lack of transferring the non-linguistic skills involved in publishing. This could prove to be an effective area where higher education policy can develop interventions.

What these hypotheses show together is that intra-national inequalities constitute the scientific lingua franca as an additional dimension of the social hierarchy in the production of knowledge in non-Anglophone countries. Early preparation in English, studying abroad, and engagement with STEM fields are the most distinguishing factors. At the national level, these have implications for policymakers in ensuring channels for quality English preparation at early stages, increasing support for studying abroad, and mitigating the foreign language deficiencies in non-STEM fields.

While the Ministry of National Education has begun foreign language-intensive education at the 5th grade level in 620 schools in 2017 (General Directorate of General Education 2017) and has increased the number of high schools with preparatory programs in the recent years, a coherent policy of foreign language education must be implemented consistently, as suggested by the cancellation of preparatory programs for certain high schools in 2005. The Ministry also implements two separate programs of scholarships and fellowships abroad, one entirely maintained by the Ministry and the other through bilateral agreements with a number of countries. These programs could benefit from

increasing quotas for countries deemed more conducive to increasing internationally mobile students integration into international science, increasing the quantitative criteria for selection and removing the categorical reliance on verbal assessment by the Ministry itself for the first program (General Directorate of Higher Education and Education Abroad n.d.), and removing the seemingly arbitrary selection criterion of a candidate's "Capability for Representation" for the bilateral program (General Directorate of European Union and Foreign Relations 2022).²²

It is also possible that such attempts might exacerbate intra-national inequalities through social reproduction, as demonstrated by this study, due to which special attention should be paid by universities and YÖK to increasing resources in academic writing for researchers. Recent developments in machine learning also show that linguistic support services can now be provided at much lower costs, suggesting that it is an opportune moment to develop supportive services in higher education organizations.

Directions for Future Research

There is much room for exploration and improvement of the models I have put forth for the field. It is likely that the chronologically earlier variables have larger consequences through indirect effects on English publications, as well as other related dependent variables. Researchers can establish this through the analysis of multiple dependent variables. Possible direct and indirect effects might also be observed through mediation analysis, which is absent in this study. Furthermore, future research must look

²² A fair assessment of state policies in this regard requires more space and a systematic review.

at the organizational and departmental variables as independent as well as dependent/moderator variables, as they can both function as social resources that further differentiate research output (e.g., some universities providing better resources and an environment for research) and as outcomes that reflect earlier differences in resources and demographic characteristics.

Different types of data (such as more information related participants' high schools, the regional origin of bachelor's degrees, and parental occupations) could reveal that the pre-PhD period has a larger impact that explored in this study. Other variables that have been explored here could be better analyzed through disaggregating the grouped variables about ethnicities, departments, and regional origins. Furthermore, corpus analysts can try to regress with similar variables on bibliometric data instead of approximate percentages reported by participants.

An important need is to combine such findings with qualitative research, a necessity in elaborating the social mechanisms behind causal relations. Separately, this is also important to test whether the suggestion that graduate-level Turkish higher education has deficiencies in transferring non-linguistic publication skills, as policymaking cannot solely rely on the implications of quantitative analysis.

REFERENCES

- Akçiğit, Ufuk, and Elif Özcan-Tok. 2020. *Türkiye Bilim Raporu [The Turkey Science Report]*. Ankara: Turkish Academy of Sciences.
- Allen, Michael Patrick. 1997. *Understanding Regression Analysis*. New York and London: Plenum Press.
- Altbach, Philip G., and Jamil Salmi, eds. 2011. *The Road to Academic Excellence*. The World Bank.
- Ammon, Ulrich. 2001. "English as a Future Language of Teaching at German Universities? A Question of Difficult Consequences, Posed by the Decline of German as a Language of Science." Pp. 343–61 in *The Dominance of English as a Language of Science*, edited by U. Ammon. Berlin and New York: De Gruyter Mouton.
- Arkin, Erkan, and Necdet Osam. 2015. "English-Medium Higher Education: A Case Study in a Turkish University Context." Pp. 177–200 in *English-Medium Instruction in European Higher Education*, edited by S. Dimova, A. K. Hultgren, and C. Jensen. De Gruyter Mouton.
- Bocanegra-Valle, Ana. 2015. "Peer Reviewers' Recommendations for Language Improvement in Research Writing." Pp. 207–30 in *English as a Scientific and Research Language*, edited by R. Plo Alastrué and C. Pérez-Llantada. De Gruyter Mouton.
- Bourdieu, Pierre. 1977. "The Economics of Linguistic Exchange." *Social Science Information* 16(6):645–68.
- Bourdieu, Pierre. 1984. *Distinction: A Social Critique of the Judgement of Taste*. Cambridge, Massachusetts: Harvard University Press.
- Bourdieu, Pierre. 1991. *Language and Symbolic Power*. Cambridge: Harvard University Press.
- Breeze, Ruth. 2015. "Citing Outside the Community? An Investigation of the Language of Bibliography in Top Journals." Pp. 37–58 in *English as a Scientific and Research Language*, edited by R. Plo Alastrué and C. Pérez-Llantada. De Gruyter Mouton.
- Bush, Vannevar. 1945. *Science, the Endless Frontier: A Report to the President*. Washington: U.S. Govt. Print. Off.
- Council of Higher Education. 2023. "Universities." Retrieved December 7, 2023 (<https://akademik.yok.gov.tr/AkademikArama/view/universityListview.jsp>).

- Council of Higher Education. n.d. “Atanma Kriterleri (Appointment Criteria).” Retrieved December 15, 2023 (<https://www.yok.gov.tr/en/academic/appointment-criteria>).
- David, Crystal. 2003. *English as a Global Language*. 2nd ed. Cambridge: Cambridge University Press.
- Davison, Roderic H. 1994. “On Dokuzuncu Yüzyıl Osmanlı Reformlarına Araç Olarak Fransız Dili [The French Language as a Vehicle for Ottoman Reform in the Nineteenth Century].” *Çağdaş Türkiye Tarihi Araştırmaları Dergisi* 2(4):171–88.
- Dontcheva-Navratilova, Olga. 2015. “Cross-Cultural Variation in Citation Practices: A Comparative Analysis of Citations in Czech English-Medium and International English-Medium Linguistics Journals.” Pp. 185–206 in *English as a Scientific and Research Language*, edited by R. Plo Alastrué and C. Pérez-Llantada. De Gruyter Mouton.
- Education First. 2023. “EF English Proficiency Index 2023.” Retrieved December 9, 2023 (<https://www.ef.com/assetscdn/WIBIwq6RdJvcD9bc8RMd/cefcom-epi-site/reports/2023/ef-epi-2023-english.pdf>).
- General Directorate of European Union and Foreign Relations. 2022. “Yabancı Hükümet Bursları Genel Şartlar (General Requirements for Foreign Government Scholarships and Fellowships).” Retrieved December 11, 2023 (<https://abdigm.meb.gov.tr/www/yabanci-hukümet-burslari-genel-sartlar/icerik/246>).
- General Directorate of General Education. 2017. “Ortaokul 5. Sınıflarda Yabancı Dil Ağırlıklı Eğitim Uygulaması [Implementing Foreign Language-Intensive Education for Middle Schools at the 5th Grade].” Retrieved December 11, 2023 (<https://tegm.meb.gov.tr/www/ortaokul-5-siniflarda-yabanci-dil-agirlikli-egitim-uygulamasi/icerik/471>).
- General Directorate of Higher Education and Education Abroad. n.d. “Sıkça Sorulan Sorular (Frequently Asked Questions).” Retrieved December 11, 2023 (<https://yyegm.meb.gov.tr/www/sss.php>).
- Gnutzmann, Claus, Jenny Jakisch, and Frank Rabe. 2015. “Resources for Publishing in English as a Foreign Language: Strategies, Peers and Techniques.” Pp. 59–84 in *English as a Scientific and Research Language*, edited by R. Plo Alastrué and C. Pérez-Llantada. De Gruyter Mouton.
- Hacettepe University Institute of Population Studies. 2015. *2013 Türkiye Nüfus ve Sağlık Araştırması İleri Analiz Çalışması [2013 Demographic and Health Survey of Turkey, Advanced Analysis Report]*. Ankara: Hacettepe University Institute of Population Studies, Ministry of Development, and TÜBİTAK.

- Hewings, Ann, Theresa Lillis, and Dimitra Vladimirova. 2010. "Who's Citing Whose Writings? A Corpus Based Study of Citations as Interpersonal Resource in English Medium National and English Medium International Journals." *Journal of English for Academic Purposes* 9(2):102–15.
- International Labour Organization. 2008. "International Standard Classification of Occupations (ISCO)." Retrieved December 8, 2023 (<https://ilostat.ilo.org/resources/concepts-and-definitions/classification-occupation/>).
- Kaplan, Robert B. 2011. "English – the Accidental Language of Science?" Pp. 3–26 in *The Dominance of English as a Language of Science*, edited by U. Ammon. Berlin and New York: De Gruyter Mouton.
- Koru, Selim, and Jasper Akesson. 2011. *Turkey's English Deficit*. Ankara: The Economic Policy Research Foundation of Turkey (TEPAV).
- Kryuchkova, Tatjana. 2001. "English as a Language of Science in Russia." Pp. 405–23 in *The Dominance of English as a Language of Science*, edited by U. Ammon. Berlin and New York: De Gruyter Mouton.
- Kuşçu, Ertan, and Ali Yağlı. 2022. "Yabancı Dil Olarak Fransızcanın Osmanlı Devleti'nin Son Dönemlerindeki Siyasal, Kültürel ve Eğitsel Alanlardaki Hâkimiyeti." *Kafkas University Journal of the Institute of Social Sciences* 30:467–88.
- Kuteeva, Maria. 2015. "Academic English as 'Nobody's Land': The Research and Publication Practices of Swedish Academics." Pp. 261–80 in *English as a Scientific and Research Language*, edited by R. Plo Alastrué and C. Pérez-Llantada. De Gruyter Mouton.
- Kuteeva, Maria, and Lisa McGrath. 2014. "Taming Tyrannosaurus Rex: English Use in the Research and Publication Practices of Humanities Scholars in Sweden." *Multilingua - Journal of Cross-Cultural and Interlanguage Communication* 33(3–4):365–87. doi: 10.1515/multi-2014-0017.
- Lillis, Theresa, and Mary Jane Curry. 2010. *Academic Writing in a Global Context: The Politics and Practices of Publishing in English*. London and New York: Routledge.
- Marks, Gary Neil. 2008. "Are Father's or Mother's Socioeconomic Characteristics More Important Influences on Student Performance? Recent International Evidence." *Social Indicators Research* 85(2):293–309.

- McGrath, Lisa. 2014. "Parallel Language Use in Academic and Outreach Publication: A Case Study of Policy and Practice." *Journal of English for Academic Purposes* 13:5–16. doi: 10.1016/j.jeap.2013.10.008.
- Medgyes, Péter, and Robert B. Kaplan. 1992. "Discourse in a Foreign Language: The Example of Hungarian Scholars." *International Journal of the Sociology of Language* 98:67–100.
- Medgyes, Péter, and Mónika László. 2011. "The Foreign Language Competence of Hungarian Scholars: Ten Years Later." Pp. 261–86 in *The Dominance of English as a Language of Science*, edited by U. Ammon. Berlin and New York: De Gruyter Mouton.
- Montgomery, Scott L. 2013. *Does Science Need a Global Language? English and the Future of Research*. Chicago and London: the university of chicago press.
- National Center for Education Statistics. 2020. "The Classification of Instructional Programs." Retrieved December 8, 2023 (<https://nces.ed.gov/ipeds/cipcode/default.aspx?y=55>).
- Oliver, Sonia. 2015. "Spanish Authors Dealing with Hedging or the Challenges of Scholarly Publication in English L2." Pp. 141–58 in *English as a Scientific and Research Language*, edited by R. Plo Alastrué and C. Pérez-Llantada. De Gruyter Mouton.
- Özkan, Selim Hilmi. 2010. "Osmanlı Devleti'nde Yabancı Dil Eğitimi." *Turkish Studies - International Periodical For the Languages, Literature and History of Turkish or Tur-Kic* 5(3):1783–1800.
- Plo Alastrué, Ramón. 2015. "Debates and Discourses on English as an Academic and Research Language." Pp. 3–8 in *English as a Scientific and Research Language*, edited by R. Plo Alastrué and C. Pérez-Llantada. De Gruyter Mouton.
- Roll-Hansen, Nils. 2017. "A Historical Perspective on the Distinction Between Basic and Applied Science." *Journal for General Philosophy of Science* 48:535–51.
- Salmi, Jamil. 2009. *The Challenge of Establishing World Class Universities*. The World Bank.
- Schluer, Jennifer. 2015. "English as a Lingua Franca in Linguistics? A Case Study of German Linguists' Language Use in Publications." Pp. 233–60 in *English as a Scientific and Research Language*, edited by R. Plo Alastrué and C. Pérez-Llantada. De Gruyter Mouton.

- Scimago Lab. 2022. "Country Rankings." Retrieved December 9, 2023 (<https://www.scimagojr.com/countryrank.php>).
- Selvi, Ali Fuad. 2011. "World Englishes in the Turkish Sociolinguistic Context." *World Englishes* 30(2):182–99.
- Starfield, Sue. 2001. "'I'll Go with the Group': Rethinking 'Discourse Community' in EAP." Pp. 132–47 in *Research Perspectives on English for Academic Purposes*, edited by J. Flowerdew and M. Peacock. Cambridge: Cambridge University Press.
- Swales, John M. 1990. *Genre Analysis: English in Academic and Research Settings*. Cambridge: Cambridge University Press.
- Swales, John M. 1997. "English as Tyrannosaurus Rex." *World English* 16(3):373–82.
- Swales, John M. 2001. "EAP-Related Linguistic Research: An Intellectual History." Pp. 42–54 in *Research Perspectives on English for Academic Purposes*, edited by J. Flowerdew and M. Peacock. Cambridge: Cambridge University Press.
- Tillé, Yves, and Alina Matei. 2016. "Basics of Sampling for Survey Research." in *The SAGE Handbook of Survey Methodology*, edited by C. Wolf, D. Joye, and T. W. Smith. SAGE Publications Ltd.
- Times Higher Education. 2022. "World University Rankings 2023." Retrieved December 7, 2023 (<https://www.timeshighereducation.com/world-university-rankings/2023/world-ranking>).
- TurkStat. 2023. "Women in Statistics, 2022." Retrieved (<https://data.tuik.gov.tr/Bulten/Index?p=%C4%B0statistiklerle-Kad%C4%B1n-2022-49668&dil=2>).
- URAP Research Laboratory. 2022. "World Ranking." Retrieved December 7, 2023 (https://urapcenter.org/Rankings/2022-2023/World_Ranking_2022-2023).
- Wissmann, Malte, Helge Toutenburg, and Shalabh. 2007. *Role of Categorical Variables in Multicollinearity in the Linear Regression Model. Technical Report*. 008. University of Munich Department of Statistics.
- Yeğen, Mesut. 2004. "Citizenship and Ethnicity in Turkey." *Middle Eastern Studies* 40(6):51–66.
- Yurdakul, Serkan, and Sümeyye Ayyüce Şahin Demir. 2022. *Yükseköğretime Bakış 2022: İzleme ve Değerlendirme Raporu [A Glance at Higher Education 2022: The Assessment Report]*. Ankara: Eğitim-Bir-Sen.