Language of citation and publishing performance of graduate students in French-speaking countries with different economic and linguistic advantages

Rassim Khelifa\textsuperscript{a,b}, Hayat Mahdjoub\textsuperscript{c}, Aeffe Baaloudj\textsuperscript{d} and Sara Chaib\textsuperscript{d}

\textsuperscript{a}Zoology Department, University of British Columbia, Vancouver, BC V6T 1Z4, Canada; \textsuperscript{b}Biodiversity Research Centre, University of British Columbia, Vancouver, BC V6T 1Z4, Canada; \textsuperscript{c}Department of Evolutionary Biology and Environmental Science, University of Zurich, Zurich, Switzerland; \textsuperscript{d}Laboratory LBEE: Biology, Water and Environment, Faculty SNV-STU, University 8 May 1945 Guelma, Guelma, Algeria

\*rassimkhelifa@gmail.com

Abstract

The performance of graduate students in research varies greatly across countries due to various factors, mainly socioeconomic and linguistic. The current situation is critical because the wealthiest countries are also the most linguistically equipped to navigate the English-dominant landscape of academia. Here, we assess the language of citations and the publishing performance of graduate students from three French-speaking countries: Algeria, Canada, and France, where Algeria is the least English proficient and the most economically disadvantaged. We found that the bibliography of PhD theses were English dominated in all regions (72.5\% in Algeria compared with >93.1\% in Western countries), whereas those of Masters theses were French dominated in Algeria (63.3\%), relatively bilingual in France (47.6\% French), but English dominated in Canada-Québec (94.7\%) and Canada-BC (98.7\%). Algerian PhD students produced fewer papers, were less likely to publish in journals with calculated impact factors, and received fewer citations than students who graduated from universities in France or in two Canadian provinces, British Columbia and Québec. Our results suggest that the economic and linguistic disadvantages faced by graduate students from non-Western backgrounds affect their academic performance, highlighting important issues in facing future global challenges.
Résumé

La performance académique des étudiants varie considérablement d’un pays à l’autre en raison de divers facteurs, principalement socio-économiques et linguistiques. La situation actuelle est critique car les pays les plus riches sont également les mieux équipés linguistiquement pour naviguer le paysage académique dominé par la langue anglaise. Dans cet article, nous évaluons la langue des citations et les performances de publication des étudiants de trois pays francophones : l’Algérie, le Canada et la France, où l’Algérie est la moins anglophone et la plus défavorisée économiquement. Nous avons constaté que la bibliographie des thèses de doctorat était à dominance anglaise dans toutes les régions (72,5% en Algérie contre >93,1% dans les pays occidentaux) alors que celle des thèses de master était à dominance française en Algérie (63,3%), relativement bilingue en France (47,6% français) mais l’anglais domine au Canada-Québec (94,7 %) et au Canada-BC (98,7 %). Les doctorants algériens ont produit moins d’articles, étaient moins susceptibles de publier dans des revues avec des facteurs d’impact calculés et ont reçu moins de citations que les étudiants diplômés d’universités en France ou dans les deux provinces canadiennes, la Colombie-Britannique et le Québec. Nos résultats suggèrent que les désavantages économiques et linguistiques auxquels sont confrontés les étudiants diplômés issus de milieux non occidentaux affectent leurs performances académiques, mettant en évidence des problèmes importants pour faire face aux futurs défis mondiaux.

Mots-clés: publication, université, biologie, doctorat, master, éducation, science

Introduction

Students and researchers from the Global North are more successful in academia than those from the Global South due to clear economic and linguistic advantages (Amano and Sutherland 2013; Das et al. 2013), as well as other historical factors (Trisos et al. 2021). While economic and linguistic factors individually create important barriers for researchers to survive and succeed in academia, the current situation is such that the wealthiest countries, mostly from the Global North, are also the most linguistically equipped to perform and communicate research (O’Neil 2018). Researchers from the Global South thus face two gigantic challenges: performing research with limited resources (e.g., access to journals, laboratory equipment, and funding) and learning a new language to have access to science and share research in the “international scientific scene” (Núñez et al. 2019). This discrepancy in the potential of researchers to excel and reach leadership positions in academia is a huge impediment to addressing planetary challenges that require global solutions (Núñez et al. 2021).

English is the most used language in science (Montgomery 2004; Englander and Corcoran 2019) and is a requirement for publication in major journals in most fields. As a consequence, native English-speaking students and researchers have clear advantages in writing, reading, and comprehending literature in their subject and, hence, have easier access to scientific content. Students and scientists speaking English as a foreign language (EFL), on the other hand, must spend time overcoming this barrier and do not typically find the necessary infrastructure to remediate it (Ramírez-Castañeda 2020). Ultimately, this means that EFL researchers experience additional bottlenecks before they can communicate their research internationally and be recognized by the scientific community (Amano et al. 2016). Furthermore, studies have recently highlighted the weakness of a monolingual academic system (Konno et al. 2020; Núñez and Amano 2021). For example, the fact that researchers from the Global North typically cite English literature and ignore...
research written in other languages might bias the results of large meta-analyses and potentially mislead decision-making.

By understanding what factors correlate with language barriers and scientific success, new solutions may be developed to begin to address identified gaps (Amano et al. 2016; Nuñez et al. 2019; Tseng et al. 2020). Here, we conducted a study on the citation pattern of masters (MSc) and doctorate (PhD) students in biological sciences from universities in three French-speaking countries (Algeria, Canada, and France) where English is the first (English part of Canada), second (French part of Canada and France), or third language (Algeria) (Fig. 1). We specifically analyzed the frequency of citation of English literature compared with French and other languages. We selected a set of universities from each region that have biological sciences and provide online access to theses (Table 1). For each university and degree, we randomly selected a set of theses and recorded the number of references cited in each, while also individually categorizing the language of each reference (English, French, and other). We then assessed the publishing performance of 20 students from two universities in each of the four regions by assessing the number of papers published, the occurrence of an impact factor in the journal selected for publication, and the total number of citations that the publications received. We finally tested for the relationship between the proportion of English citations in a thesis and the publishing performance of PhD students.

Fig. 1. The level of English proficiency of Algeria compared with two other French-speaking countries. Countries/regions were ranked from the least to the most proficient in English: Algeria, France, Canada-Québec, and Canada-British Columbia. English is the third language taught at school in Algeria, second in France, second (bilingual) in Canada-Québec, and first in Canada-British Columbia.
Methods

Data collection

We selected three French-speaking countries (Algeria, Canada, and France) that have different levels of English proficiency. Because Canada has both French- and English-speaking parts, we selected one French-speaking (Québec (QC)) and one English-speaking province (British Columbia (BC)) and treated them separately in our analyses. Ordering the regions from the most to the least proficient in English, we have Canada-BC, Canada-QC, France, and Algeria. Based on the EF English Proficiency Index, France was ranked 28th (high proficiency category) while Algeria was ranked 81st (very low proficiency category) out of 100 countries in 2020 (ef.com/wwen/epi/). In addition to English proficiency, Canada and France are much wealthier than Algeria in gross domestic product per capita rankings (World Bank 2020).

We randomly selected a set of higher education institutions (universities and research centers) that offer MSc and (or) PhD programs in biological sciences. We concentrated on institutions that make theses open access, verified by checking the website of each institution. For each degree, we selected five universities from Algeria and two universities from France, Canada-QC, and Canada-BC. We used a larger sample size for Algerian universities to assess whether French dominance in student theses is a sampling bias or a widespread phenomenon in Algeria; we likewise sampled institutions across different Algerian provinces. We accessed the MSc and PhD theses from university websites and selected theses in biological sciences that were defended between 2006 and 2021. In many cases, the number of theses per university was low and we used all the theses that we found. Otherwise we selected the top 30 theses that were shown at the top of the list generated by the platform (Table 1). To understand the language representation of English in the MSc and PhD theses, we analyzed the

Table 1. Sample size of theses published in four French-speaking regions used in the study.

<table>
<thead>
<tr>
<th>Region</th>
<th>Province</th>
<th>Latitude</th>
<th>Longitude</th>
<th>University</th>
<th>Master</th>
<th>PhD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Algeria</td>
<td>Tlemcen</td>
<td>34.91973</td>
<td>−1.30185</td>
<td>Université Abou Bakr Belkaïd Tlemcen</td>
<td>0</td>
<td>28</td>
</tr>
<tr>
<td>Algeria</td>
<td>Bouira</td>
<td>36.38199</td>
<td>3.89584</td>
<td>Université Akli Mohand Oulhadj – Bouira</td>
<td>30</td>
<td>0</td>
</tr>
<tr>
<td>Algeria</td>
<td>Annaba</td>
<td>36.91253</td>
<td>7.74884</td>
<td>Université Badji Mokhtar Annaba</td>
<td>26</td>
<td>26</td>
</tr>
<tr>
<td>Algeria</td>
<td>Ouargla</td>
<td>31.94816</td>
<td>5.2684</td>
<td>Université Kasdi Merbeh Ouargla</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>Algeria</td>
<td>Jijel</td>
<td>36.80126</td>
<td>5.75259</td>
<td>Université Mohamed Seddik BenYahia, Jijel</td>
<td>0</td>
<td>14</td>
</tr>
<tr>
<td>Algeria</td>
<td>Tizi-Ouzou</td>
<td>36.69806</td>
<td>4.05812</td>
<td>Université Mouloud Mammeri de Tizi-Ouzou</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>Canada_BC</td>
<td>British Columbia</td>
<td>49.27802</td>
<td>−122.91982</td>
<td>Simon Fraser University</td>
<td>25</td>
<td>23</td>
</tr>
<tr>
<td>Canada_BC</td>
<td>British Columbia</td>
<td>49.26045</td>
<td>−123.24624</td>
<td>The University of British Columbia</td>
<td>30</td>
<td>25</td>
</tr>
<tr>
<td>Canada_QC</td>
<td>Québec</td>
<td>45.50561</td>
<td>−73.61379</td>
<td>Université de Montréal</td>
<td>29</td>
<td>30</td>
</tr>
<tr>
<td>Canada_QC</td>
<td>Québec</td>
<td>45.5126</td>
<td>−73.5605</td>
<td>Université du Québec à Montréal</td>
<td>0</td>
<td>22</td>
</tr>
<tr>
<td>France</td>
<td>Rennes</td>
<td>48.1137</td>
<td>−1.70619</td>
<td>Agrocampus Ouest</td>
<td>20</td>
<td>0</td>
</tr>
<tr>
<td>France</td>
<td>Montpellier</td>
<td>43.61728</td>
<td>3.85507</td>
<td>SupAgro Montpellier</td>
<td>24</td>
<td>0</td>
</tr>
<tr>
<td>France</td>
<td>Bordeaux</td>
<td>44.8012</td>
<td>−0.61013</td>
<td>Université de Bordeaux</td>
<td>0</td>
<td>20</td>
</tr>
<tr>
<td>France</td>
<td>Montpellier</td>
<td>43.63399</td>
<td>3.84892</td>
<td>University of Montpellier</td>
<td>0</td>
<td>30</td>
</tr>
</tbody>
</table>

Khelifa et al.
bibliography section of each thesis, counted the number of references, and classified the language of each reference as English, French, and other. Thus for each thesis, we were able to calculate the proportion of English, French, and other languages in its bibliography. Supplementary Table S1 presents the mean ± SD of the number of references for each university and region.

To test the potential correlation between the proportion of English citations and the scientific productivity and success of students, we randomly selected 20 PhD students from two universities of each region and counted the number of first-authored papers (related to the PhD thesis) published in scientific journals. Because each thesis had an identification number, the random selection was carried out using Excel. We kept the sample size of universities (N = 2) and students (N = 20) constant across regions to avoid any bias. Here we did not include MSc theses because unlike PhD students who must publish at least a paper to defend their theses, MSc student do not have to publish (and typically do not). Thus, the publishing performance of students was assessed only on PhD students. In most cases, the author listed the papers published or planned to be published in the thesis (we considered only those published by the time of our search by cross-checking with Google and Google Scholar). Otherwise, the author’s name was searched on Google and Google Scholar, and the first-authored publications were checked if they included the results showed in the thesis. We counted publications derived from the thesis that were published during and after the defense. Using 2019 InCites Journal Citation Reports from Clarivate, we determined whether the journal where students published their research has an impact factor (hereafter impact factor journal). Book chapters were excluded from the analysis of impact factors. Using Google Scholar, we collected the number of citations to each publication on 25 May 2021 and calculated the number of citations per year by dividing the total number of citations by the number of years since publication (years have decimal number as they were calculated to the nearest day). The three metrics of scientific productivity and success of each student (number of publications, probability of publishing in an impact factor journal, and number of citations per year) were used as response variables to assess their correlation with the proportion of English citations in their PhD thesis.

Statistical analyses

We conducted all analyses using R 4.0.2 (R Development Core Team 2021). All mixed models were computed using the R package lme4 (Bates et al. 2015). To determine potential differences between the frequency of English citations in MSc and PhD theses and across regions, we used a logistic mixed-effects model that includes the proportion of English among all citations, as a response variable, with “degree” and “region” as fixed effects. We tested for the difference in the publishing performance (total number of publications, occurrence of an impact factor of the journal selected for publication (presence/absence), and the number of citations received) among regions by using generalized mixed-effects models (GLMM) that include student (except for number of publications) and university as random effects, region as main effect, and error distribution that depends on the response variable (Poisson for the number of publications and citations and binomial for the probability of publishing in an impact factor journal). Similarly, we used GLMM to test for the correlation between the proportion of English citations in a PhD thesis and the three variables of publishing performance, including the proportion of English citations as a main effect and student as random effects. GLMMs analyzing the number of publications included year of defense as a covariate to correct for the fact that authors who defended their thesis more recently might have published less than those who defended years ago, whereas those analyzing the number of citations included as a covariate the number of days since the publications to a baseline date (1 June 2021) to correct for the fact that papers published more recently usually receive less citations than those published years ago. We added an observation-level random effect for all Poisson GLMM to account for overdispersion (Harrison 2014).
Results

Based on 299 MSc and 278 PhD students who graduated between 2006 and 2021, our results showed that students who graduated from Algerian universities cited less English literature than those who graduated from France ($z = 3.08$, $P = 0.0002$), Canada-QC ($z = 8.34$, $P < 0.0001$) and Canada-BC ($z = 11.71$, $P < 0.0001$), and this pattern was more pronounced in MSc than PhD theses ($z = 8.70$, $P < 0.0001$) (Fig. 2). The least to the most English-dominant citations for MSc are Algeria (35.9%), France (48.1%), Canada-QC (94.8%), and Canada-BC (98.6%) and similarly for PhD, Algeria (70.3%), France (93.2%), Canada-QC (96.5%), and Canada-BC (99.6%) (Supplementary Table S2).

There was a significant difference among regions in publishing performance of PhD students where students graduating from France, Canada-QC, and Canada-BC showed higher number of publications published during their PhD ($z = 2.94$, $P = 0.003$; $z = 2.70$, $P = 0.007$; $z = 2.89$, $P = 0.003$; respectively), higher probability of publishing in an impact factor journal ($z = 5.20$, $P < 0.0001$; $z = 5.84$, $P < 0.0001$; $z = 6.47$, $P < 0.0001$; respectively), and number of citations that the publications received (corrected for the time since publication) ($z = 6.98$, $P < 0.0001$; $z = 6.79$, $P < 0.0001$; $z = 6.60$, $P < 0.0001$; respectively) than Algerian students (Fig. 3).

We found a positive correlation among the proportion of English citations in a PhD thesis, the total number of papers published ($z = 2.98$, $P = 0.002$), the probability of publishing in an impact factor journal ($z = 4.54$, $P < 0.0001$), and the number of citations received since the publication ($z = 5.72$, $P < 0.0001$) (Fig. 4). These correlations were mainly driven by the geographic variation in the proportion of English citations and academic performance across the four regions; specifically, students at Algerian universities had the least English-dominant literature and had a lower likelihood of publishing in an impact factor journal and their papers received fewer citations compared with those who graduated from French and Canadian institutions.

Fig. 2. Proportion of the language of citation (English, French, and other) in the reference section of Masters and PhD students in Algeria, France, French-speaking Canada (Québec), and English-speaking Canada (British Columbia).
Discussion

Our study found that students from different French-speaking countries with different levels of English proficiency showed different tendencies of citing English literature; students from universities in more English-proficient places cited dominantly English literature, whereas those from less English-proficient places showed a weaker use of English. Moreover, the more English-proficient the area, the higher the students' performance in publishing research.

Given that English dominates the literature, an English-dominant reference list is generally the norm. The observed deviation of this norm in Algeria could be explained by a few nonmutually exclusive hypotheses. First, Algerian students typically have low English proficiency, which might force them...
to read French documentation and avoid English literature. This is particularly true for Algerian MSc students who are given only one semester to perform their project and write and defend their theses and thus do not have enough time to struggle with understanding English documents. This low English proficiency is mainly due to the quasi-absence of English in the curriculum, the lack of English training prior to graduate school, and the French-language dominance in the Algerian higher education system (Daoud 2000; Mohammed and Brahim 2010). Thus, these students probably find it easier to read, analyze, and cite studies published in more familiar languages (Guardiano et al. 2007). Studies supporting this hypothesis showed that EFL graduate students spend more time reading and understanding papers in English compared with papers written in their native language (Ramírez-Castañeda 2020). Second, due to historical reasons, it is also likely that a large proportion of Algerian literature is written in French, which restricts students to mostly cite literature in French. For instance, many of the pioneering books describing the fauna of Algeria were written by French authors in the 19th century during the so-called “Exploration Scientifique de l’Algérie” (Lucas 1849).

Although the observed geographic variation in academic performance could be partly explained by the language, other factors such as socioeconomic status, level of training (university and supervisors), and history play a crucial role (Amano and Sutherland 2013; Das et al. 2013; Nuñez et al. 2021). The fact that graduate students from France do as well as those graduating from native English-speaking countries such as Canada not only suggests that local supervisors, institutions, and academic systems can provide the adequate environment required to succeed in research, but also that English as a second official language is not as limiting as when it is the third or fourth language. Within and across regions, our results suggest that the academic environment associated with low English proficiency forces students to publish in low-profile journals that do not attract visibility and typically receive fewer citations (Nuñez et al. 2019). Ultimately, the absence of an English-dominant academic environment reduces the competitiveness of EFL students in the academic world and creates isolated scientific minorities that do not reach leadership positions in academia (Lund 2021).

The pattern observed here most likely occurs in other regions where English is a foreign language (65% of countries). This is, therefore, a global problem in science that prevents the real “open-science” objective from progressing beyond the open-access initiative. Among the top-priority solutions to solve the language barrier is the translation of scientific articles (or at least the abstract) into different languages. While publishers could endorse such a measure, the scientific community could also immediately create a sustainable system by introducing a reward system that the scientific community recognizes and values (similar to the review process) where scientists who are proficient in English and (or) other languages offer linguistic services such as language editing of draft papers written by non-native English speakers. One effective way to manage such services is to integrate them in public repository platforms such as BioRxiv and Authorea (Khelifa et al. 2021). Peer-proofing systems in preprint repositories should generate recognition metrics for contributors, similar to the peer-review metric points in Publons. Another solution might be that journals could incorporate a preliminary linguistic revision prior to the typical scientific peer-review for EFL scientists and such revisions should be recognized as linguistic peer-review by Publons. Such measures will solve a large problem in science accessibility and citation of studies published in different languages and, hence, may reduce the bias of unilingual literature reviews (Konno et al. 2020), thereby helping EFL students publish in high-impact journals that increase the visibility of scientific minorities (Lebel and McLean 2018; Nuñez et al. 2019). Some journals have started tackling the language barrier using different initiatives. For instance, journals such as Journal of Agriculture, Food Systems, Community Development and Human-Wildlife Interactions have already established a language assisting system for authors who are not proficient in English. European Journal of Ecology states in their homepage that the journal provides “free language-correction services for authors from non-English speaking regions”
In addition, EFL institutions and researchers should meet the scientific community halfway. Knowing that English is the lingua franca in science (O’Neil 2018), EFL institutions should renovate their strategic planning to improve the English proficiency of students well before graduate school. This might involve increasing the mandatory number of learning hours in English, introducing new courses for academic communication, promoting exchange programs with foreign institutions that teach and do research in English, and engaging in collaborations with students and researchers who are proficient in English. All these measures, including the international collaborations, are nowadays readily implemented with video and audio conferencing, particularly in response to the COVID-19 pandemic that temporarily shifted the pedagogical and academic landscape to online-based communication and e-learning in many regions around the world (Reshef et al. 2020).

The exorbitant disparity in academic success that currently exists could be reduced if the scientific community, research institutions, publishers, funders, and governments take part in undertaking actionable measures to overcome the language barrier, socioeconomic differences, cultural privileges, and discrimination. This is crucial for the global inclusion of under-represented scientists in academia and the diversification of opinions and solutions for contemporary planetary problems such as climate change, biodiversity loss, and disease epidemics (Amano and Sutherland 2013; Angulo et al. 2021). An increasing number of calls from the scientific community to address issues of diversity and inclusion in science have been observed during the last few years. Less commonly noted, but certainly urgent, is the need for coordinated measures and strategic commitment to improve knowledge transfer and balance the representation of EFL researchers in the rapidly growing body of scientific literature.

Acknowledgements
RK is funded by the Swiss National Science Foundation (project: P400PB_191139). We thank two anonymous reviewers and the Editor for their helpful comments and suggestions, which improved the quality of an earlier version of the manuscript. We are grateful to Leithen K. M’Gonigle for his fruitful comments on the manuscript.

Author contributions
RK conceived the presented idea. RK and HM designed the survey study. RK, HM, AB, and SC collected the data. RK and HM performed the analysis and wrote the manuscript with input from AB and SC. RK, HM, AB, and SC gave final approval for publication.

Competing interests
The authors declare no competing interests.

Data availability statement
All relevant data are within the paper and in the Supplementary Material.
Supplementary materials

The following Supplementary Material is available with the article through the journal website at doi:10.1139/facets-2021-0073.

Supplementary Material 1

Supplementary Material 2

References

Amano T, González-Varo JP, and Sutherland WJ. 2016. Languages are still a major barrier to global science. PLoS Biology, 14: e2000933. PMID: 28033326 DOI: 10.1371/journal.pbio.2000933


