CARTA INTERNACIONAL
ASSOCIAÇÃO BRASILEIRA DE RELAÇÕES INTERNACIONAIS

ISSN 2526-9038

The determinants of inbound international tertiary students in the developing world: the global south dimension

Os determinantes da entrada de estudantes internacionais de ensino superior no mundo em desenvolvimento: a dimensão do sul global

Los determinantes del ingreso de estudiantes internacionales de educación superior en el mundo en desarrollo: la dimensión del sur global

DOI: 10.21530/ci.v19n2.2024.1517

Rafael de Moraes Baldrighi[1]

Abstract

Over recent decades, the internationalization of universities has become a global norm. Tertiary student mobility literature identifies push-pull factors driving cross-border flows, but does it explain why developed countries dominate as destinations while emerging ones, the Global South, export students? Using Ordinary Least Squares multiple regression, we analyzed pull factors influencing inbound students across emerging countries. Our findings highlight the significant role of geopolitical factors, with language, academic excellence, migrant networks, and hosting capacity also positively affecting student inflows.

Keywords: Student Mobility; Academic Internationalization, Global South; Higher Education.

1 Master's Degree in International Relations at the Institute of International Relations (IRI) – University of São Paulo (USP). Rafael is a public servant and researcher for the Brazilian Government at the Institute for Applied Economic Research (Ipea). (rafaelbaldrighi@gmail.com). ORCID: **https://orcid.org/0000-0002-7527-3404**.

Artigo submetido em 16/08/2024 e aprovado em 02/12/2024.

Resumo

Nas últimas décadas, a internacionalização das universidades tornou-se uma norma global. A literatura sobre mobilidade estudantil no ensino superior identifica fatores de atração e expulsão que impulsionam os fluxos transfronteiriços, mas será que isso explica por que os países desenvolvidos predominam como destinos, enquanto os emergentes, do Sul Global, exportam estudantes? Utilizando regressão múltipla por mínimos quadrados ordinários, analisamos os fatores de atração que influenciam os estudantes internacionais em países emergentes. Nossos resultados destacam o papel significativo de fatores geopolíticos, além do impacto positivo da língua, excelência acadêmica, redes migratórias e capacidade de acolhimento.

Palavras-chave: Mobilidade Estudantil; Internacionalização Acadêmica; Sul Global; Ensino Superior.

Resumen

En las últimas décadas, la internacionalización de las universidades se ha consolidado como una norma global. La literatura sobre movilidad estudiantil terciaria analiza factores de atracción y expulsión que impulsan los flujos transfronterizos, pero ¿explica por qué los países desarrollados son destinos principales mientras que los emergentes, el Sur Global, exportan estudiantes? Usando regresión por mínimos cuadrados ordinarios, analizamos los factores de atracción de estudiantes en países emergentes. Nuestros hallazgos destacan la relevancia de factores geopolíticos, así como el impacto positivo del idioma, la excelencia académica, las redes migratorias y la capacidad de acogida.

Palabras clabe: Movilidad Estudiantil; Internacionalización Académica; Sur Global; Educación Superior.

Introduction

For the last three decades, the internationalization of Higher Education Institutions (HEIs) has become a consolidated, and almost obligatory, process for well-ranked universities. Mostly inspired by the European experience, HEIs all around the world are developing internationalization strategic plans and policies, with the phenomenon of tertiary mobility, both inbound and outbound, as one of their main components. According to the Organization for Economic

Co-operation and Development (OECD), between 1998 and 2018, the total number of international tertiary students has grown by 4.8% on average, yearly, reaching almost 6 million in 2018 (OECD 2019).

However, those values are significantly imbalanced. In 2018, students from emerging countries represented the bulk of the outgoing flows migrating toward developed countries, with Asia accounting for 57% of all mobile students in OECD states, and China and India alone representing more than 30% of that number. The United States (US), on the other hand, received 18% of the world's flow, followed by Australia and the United Kingdom (8% each). In that year, less than 30% of the world's mobile students were enrolled in non-OECD countries.

The vast literature on tertiary student mobility highlights push-pull factors that can help us better understand this imbalance in which developed countries are the main destinations and emerging countries – here, the Global South (Mignolo 2002; Docquier and Rapoport 2012; Van Bouwel and Veugelers 2013; Beck and Pidgeon 2020), that is, countries where emancipation from Western established economic and political discourse used to legitimize cultural control is unfolding – are the exporters of students. Among the push factors are the value of holding a foreign degree, national isolation (geographical or cultural), and political/economic issues (academic freedom, censorship, employment opportunities, and income expectations). Among the pull factors are the existence of well-ranked institutions, a dynamic job market, hosting capacity, geographical proximity, and a plethora of other features that a country can display that are 'attractive' to foreign students (Beine, Romain and Ragot 2014; Caruso and De Wit 2014; Didisse, Nguyen-Huu and Tran 2018).

Here, we analyze the tertiary mobility phenomenon with a focus on the Global South, given the disparity in the number of inbound international students, dealing with issues such as unequal globalization, brain-drain/brain-gain, former colonization ties, and the predominance of OECD HEIs in university rankings. We do so, since it is important to know which policies developing countries can design to attract students, and since tertiary mobility has become increasingly relevant over the last decades, both for the growing number of mobile students (OECD 2019) and the increase in monetary terms of education services (Caruso and De Wit 2014).

Therefore, this paper aims to study the pull factors, or the attractiveness, of inbound mobile students in a cross-section comprising emerging countries in the year of 2017 using Ordinary Least Squares (OLS) multiple regression

models. The dependent variable is the proportion of higher education incoming students per country relative to the total number of students studying abroad – regardless of its nature: mobility, research, whole course, undergraduate, graduate, Ph.D., post-doc, and so on. We limit the analysis to 2017 since this is the year with the highest availability of data for mobile students in emerging countries in the Unesco Institute for Statistics (UIS) database. Also, it predates the 2020-2022 global COVID pandemic that severely impacted mobility numbers. The explanatory variables are: spoken language, academic excellence, migrant network, geographical distance, income expectations, hosting capacity, and cost of living.

We opted for a cross-section since we deal with a significant lack of data on mobility for non-OECD countries (Beine, Romain and Ragot 2014; Rumbley 2012), hindering the construction of a panel. Also, the dependent variable and some explanatory factors are (quasi) constant over time, such as distance, language, and institutional features (Kahanec and Králiková 2011; Didisse, Nguyen-Huu and Tran 2018). Our innovation is the use of non-OECD countries as the sample – which is not abundant in the literature and official databases on the subject, as the next section shows – and the creation of three dummy variables to assess which dimension better represents the Global South – to try to capture a set of country-specific aspects that differentiate developing nations from developed ones (Caruso and De Wit 2014). Most studies on the subject, qualitative or quantitative, focus on European and OECD countries only.

Our findings, after running ten models, are that out of the three dummies used, the one that best encapsulates the Global South facet was the existence of a former colonial regime in a given country. Also, spoken language, academic excellence, a network of migrants, and hosting capacity are significant and positively related to our dependent variable. This paper is structured as follows: in the next section, we highlight the pull factors of tertiary mobility and present the Global South perspective. In "Descriptive Analysis and Data Presentation", we display the variables and the data collection and treatment procedures. In "Main Findings", we present the empirical application of the models and the outputs of the regressions. A final section summarizes our findings.

Literature Review: Pull Factors of Student Mobility and the Global South

Pull Factors of International Mobility

The literature on international mobility of tertiary students highlights some factors that can either attract or refrain students from a specific destination. Those are commonly divided into push-pull factors (Caruso and De Wit 2014). The former is related to national features of the country of origin that propel students to move abroad; and the latter regards the attractiveness of a state to lure international students. Although this dynamic is in decline in migration studies, it remains relevant in works on academic mobility, which is a singular phenomenon regarding migrants. Here, we do not focus on the factors 'at home' that push individuals to seek education abroad, but on how a country can attract foreign students.

But why is it relevant to woo international students? Besides from shielding the minds that would be lost due to student migration, the act of hosting highly-skilled migrants has direct and indirect gains in improvements in education, productivity, and research. Also, the externalities of receiving foreign students range from a more diverse and multicultural society in the micro level, to international cooperation and regime strengthening in the macro level – Docquier and Rapoport (2012) and Van Bouwel and Veugelers (2013). Sure, internationalization processes encompass international publications, the promotion of English-taught courses, bilateral agreements, and a plethora of other features that a university can display to advance in international cooperation. Its main perceptible component, though, is the mobility, incoming and outgoing, of students.

The first pull factor is language proximity. Beine, Romain and Ragot (2014) find that a common official language between origin and destination countries positively impacts the flow of students. This is also noted by Junqueira and Baldrighi (2020) when analyzing if widely spoken languages are more attractive to students by comparing Spanish and Portuguese. Didisse, Nguyen-Huu and Tran (2018) use a four-tier measure for language proximity (from a common official language to the similarity of different languages) and find a positive relationship for all measures. Equivalent results can be found in Kahanec and Králiková (2011) and OECD (2011) that highlight the national language of instruction as one of the most important factors to attract foreign students, with English, French, German, Spanish, and Russian being the most inviting ones.

Academic excellence, what Van Bouwel and Veugelers (2013) call 'the quality dimension', is also a pull factor that positively affects the number of mobile international students in a country. Measures of the academic impact of a country's scientific publications, expenditure per student, or the number of national universities in HEI rankings are the general proxies used to estimate this factor. Using the top 200 universities in the Shanghai Ranking – Academic Ranking of World Universities (ARWU) –, Beine, Romain and Ragot 2014 find that quality of education is a significant but moderate attractor when compared to other pull factors. Van Bouwel and Veugelers (2013) find, regardless of the indicator used (Shanghai Ranking, the Times Higher Education – THE – Ranking, and the relative impact of a country's scientific publications), a positive relationship for the flow of inbound students. Similar results can be found in Caruso and De Wit (2014), who detect a positive and significant relationship between the number of incoming students and public expenditure per student. In those rankings, there is a strong bias towards educational institutions from the Global North.

A network of similar social actors to which a student can rely on when abroad is another factor that is positively related to incoming students in the literature. The proportion of foreigners living in a country (Junqueira and Baldrighi 2020), the number of high-skilled migrants (Beine, Romain and Ragot 2014), and even internet users (Didisse, Nguyen-Huu and Tran 2018) are instrumentalized to measure such network effect. The higher the number of students in the destination country, the higher the flow of students from the same origin since the "presence of country nationals at destination tends to act as a magnet for international students" (Beine, Romain and Ragot 2014: 51). According to these authors, the existence of a migration network in the destination country can also reduce migration costs. Furthermore, surveys can also capture this phenomenon, as Mazzarol and Soutar (2000) and Pedro and Franco (2015) demonstrate, after surveying mobile students in Australia and Portugal, respectively.

Commonly, the geographical distance between countries negatively impacts the proportion of students attracted. In all six models presented by Beine, Romain and Ragot (2014), distance has a negative coefficient and is statistically significant at the 1% level. The same is true for Van Bouwel and Veugelers (2013) who use five simple gravity models. Caruso and De Wit (2014) also observe this negative impact and feature geographical distance as an economic pull factor since smaller distances generally translate into smaller transportation costs. From this, we can also infer that the existence of bordering countries, analyzed dyadically, facilitate

travel procedures and have more flexible visa policies, which is accounted for by Van Bouwel and Veugelers (2013).

Another common pull factor is the expected future income in the destination country. Caruso and De Wit (2014) find a positive, significant at 1%, relationship for the gross domestic product (GDP) per capita and incoming students. The same is true for Beine, Romain and Ragot (2014). Using gross annual wage for workers with tertiary education level, these authors also find a positive and significant at the 1% level impact on the dependent variable. Didisse, Nguyen-Huu and Tran (2018) apply economic and socio-demographic factors as proxies, such as youth unemployment and average enrolments in tertiary education, and find, respectively, a negative and a positive relationship.

Hosting capacity is another variable. A vast number of national universities, a big and dynamic job market, and several housing or funding opportunities can act as a magnet to attract and increase the number of foreign students. The most common proxy used is population. Beine, Romain and Ragot (2014) show that students are sensitive to this factor. However, the measures used differ: the authors use the total population (logged) as a proxy and compare it to the total number of students enrolled at the university of destination during a given academic year for Italy and the United Kingdom (UK). Similarly, and dyadically, Van Bouwel and Veugelers (2013) use the student population in the host and the sender to instrumentalize this positive relationship in basic gravity models.

Cost of living can also impact the number of inbound tertiary students and the literature points to a negative relationship. Caruso and De Wit (2014) proxy this factor by the current inflation change, whereas Beine, Romain and Ragot (2014) employ Numbeo's *Consumers Price Index*. Also, tuition fees are important when analyzing the costs of student migration. However, this data is, overall, not available for a good number of countries (Rumbley 2012; Didisse, Nguyen-Huu and Tran 2018) and Beine, Romain and Ragot (2014) find that, although living costs have a negative strong impact on incoming students, fees are insignificant. This happens since mobile students often benefit from stipends or fellowships to cover them. This is highly debatable, though, as Caruso and De Wit (2014) find that cost of living alone does not discourage the inflow of foreign students: it only does when combined with tuition fees. Collinearity issues in regression models can also arise when measuring the cost of living since, depending on the proxy, it is often strongly correlated with income per capita, public expenditure, or fees (Caruso and De Wit 2014).

Other factors, which receive less attention from the literature, can also be found in quantitative models and surveys, such as violence – which often include homicides, committed crimes, and other phenomena that are hard to grasp, such as political repression, xenophobia, and racism (Mazzarol and Soutar 2000; Caruso and De Wit 2014; Junqueira and Baldrighi 2020); public policies designed to attract students – which include scholarship policies, the promotion of English-taught courses, visa restrictions, HEIs' agreements, and migration opportunities, which, again, suffer from lack of data (Kahanec and Králiková 2011; Rumbley 2012); and other cultural and religious dyadic factors that may impact the bilateral flow of students.

The Global South Perspective

Besides explanatory factors, scholars often come from different perspectives and theoretical backgrounds to explain why students migrate, such as human capital theory – the mobility as an investment to grab job opportunities or to increase future income (Rosenweig 2008); a consumption choice – the search for a better education than at home (Van Bouwel and Veugelers 2013; Beine, Romain and Ragot2014); social capital theory – students being attracted to countries where they can find a similar social network in a cumulative causation process (Van Bouwel and Veugelers 2013; Pedro and Franco 2015); and, lastly, from a critical point of view on globalization and human migration (Mignolo 2002), which the literature identifies as 'the Global South perspective'. From this standpoint, migration flows originate in the so-called *periphery* with the so-called *core*, or the Global North, as the destination. The latter ensure their dominant position by retaining minds and talents from peripheral countries in a brain-drain/brain-gain cycle, mainly from former colonies to former colonial powers.

However, we do not use this perspective naively, blaming globalization only and what some authors call the *Geopolitics of Knowledge* (Mignolo 2002; Beck and Pidgeon 2020). Even though this dependent and imbalanced relation does exist, Docquier and Rapoport (2012) and Van Bouwel and Veugelers (2013) show that student mobility from the periphery to the core can create positive externalities at home – on technological, educational, and political issues – and not necessarily lead to more unequal or neocolonial (Buckner and Stein 2019) relations. Here, we list five factors and conduct two hypothesis tests to argue for the development of a model specifically designed for inbound mobile students in the Global South.

The first factor is the imbalanced flow. Besides the numbers in the introduction brought by OECD (2019), we can compare the proportion of international students per region of the world. With UIS' National Monitoring data for *inbound mobility rate*,[2] we find that, in the year 2017 (the reference year for data collection in our study), the world's average for this index is at 2.4%. In the Global South: Sub-Saharan Africa, 1.7%, the Arab States, 3.06%, Asia-Central, 2.16%, Asia-Southern, 0.16%, Asia-Eastern, 0.85%, Asia-South-Eastern, 1.07%, and Latin America and the Caribbean, 0.73%. In developed countries, the inbound rate is at 7.33% in North America and Western Europe, 3.43% in Central and Eastern Europe, 4.27% in Japan, and 21.27% in Oceania (Australia/New Zealand). That is, except for the Arab States – possibly due to the high rate of international students in the Gulf states and the Syrian and Palestinian diasporas –, all the Global South regions are below the world's average and significantly below developed countries' average.

Furthermore, as presented before, the majority of incoming students in the world are from emerging countries, creating a 'natural' flow that consolidates mobility as a movement of millions of minds leaving the Global South toward developed countries. UIS' *net flow of internationally mobile students*[3] gives us a hint that, in general, emerging countries present a deficit when calculating the difference between incoming and outgoing students, whereas OECD countries tend to register a surplus.

Another factor is academic excellence. As presented, university rankings are used as proxies to measure the quality dimension (Van Bouwel and Veugelers 2013). Among those, are the ARWU/Shanghai Ranking, THE, and Quacquarelli Symonds (QS) World University Ranking. Although these can be helpful and adequate measures of academic excellence, there is a strong predominance of North American and European universities, excluding a huge deal of universities from emerging states, making it difficult to compare Global South countries' academic excellence. Sure, ARWU has a strong bias toward China. Also, Brazil, Chile, India, and South Africa have a good number of HEIs in these rankings. However, to analyze academic excellence in the Global South, we need an inclusive criterion, contrasting with Van Bouwel and Veugelers (2013), which limits it to the top 200 in the ARWU ranking, and Didisse, Nguyen-Huu and Tran (2018), who limits it

2 According to UIS' glossary, it is the "number of students from abroad studying in a given country, expressed as a percentage of total tertiary enrolment in that country".

3 That is the difference between the number of students hosted and the number of students sent abroad.

just to ARWU's top 100. Restricting the number of HEIs analyzed to just a couple hundred would severely hinder a thorough analysis of Global South academic excellence, since few emerging countries would make it to the podiums.

The third feature that marks the mobility of students leaving emerging countries is a former colonial tie. This phenomenon is captured by Beine, Romain and Ragot (2014) and Didisse, Nguyen-Huu and Tran (2018) who show a positive and significant relationship between former colonial links and incoming students. For example, Brazilians are the largest group of international students in Portugal, Indians are the second-largest in the UK, and Moroccans and Algerians are, respectively, the first and the second major groups of inbound students in France. Therefore, a sample comprising emerging countries only (since states[4] in the Global South were, mainly, former colonies), would help us better capture the determinants of inbound mobility toward them.

The fourth point is the late evolution of internationalization in emerging countries. Not only the development of universities originated in Europe, but, in that continent, scholars have been mobile for centuries (Van Bouwel and Veugelers 2013). Furthermore, in the last three decades, European countries have engaged in highly successful projects to strengthen student mobility in the continent via stipend and fellowship programs supported by legal international obligations, standardization of credits, and multilateral commitments embodied by the *Magna Charta Universitatum* (1988) and the Bologna Declaration (1999). Institutions such as *Campus France* and the *Deutscher Akademischer Austauschdienst* (DAAD) are well known all over the world. In contrast, efforts in the Global South have been much more modest and came much after European initiatives.

Also, due to the high quality and tradition of American universities, the US has dominated the scientific and academic scenario since the mid-1900s, attracting a huge number of scholars and skilled students over the last decades. Chen and Barnett (2000) classify countries into three categories they created: a core destination of students (Western Europe, Oceania, US, and Canada); what they call the semi-periphery, such as Eastern Europe (mainly Russia);[5] and the

4 In the Global South sample, we include countries such as Russia and Turkey. Arguably they can be considered as former colonial powers, but due to the inclusion criteria explained in the next section, they made it to the emerging dataset.

5 Russia (and the USSR) is a very important destination country for mobile students and can be included as a traditional destination. However, due to the sampling criterion used in this paper – presented in the next section –, we considered the Russian Federation as an observation in the emerging dataset. The *col_pow45* dummy tries to deal with this issue, though.

periphery, which is not attractive to foreign students, such as Latin America, Africa, and Asia. Thus, path dependence and late development of internationalization policies corroborate our 'emerging' model since Global South countries are not yet ready to play in the big leagues on an equal footing.

Lastly, the fifth feature regarding Global South inbound mobile students is the lack of quantitative studies in the literature about them. Almost all of the empirical studies cited here use European and OECD countries as sample. Sure, it is undeniably true that the availability of data for those countries is higher, counting not only on UIS', but also on OECD's and Eurostat's databases – the UOE data sets (Rumbley 2012). Most of the works on the subject in non-European and emerging countries use qualitative methodologies (Junqueira and Baldrighi 2020), or assume a critical and theoretical point of view, without diving into statistical analyses (Mignolo 2002; Buckner and Stein 2019). There are studies that address some attractive variables to include developing countries, but there is no quantitative model, to date, that attempts to better understand the flow of international students in those nations. Thus, our model including emerging countries tries to remedy this lack in the literature.

To check on the properness of these five assumptions on why we should model emerging countries, we conduct a hypothesis test comparing UIS' inbound mobility rate for North America and Western Europe (NAWE) with the mean proportion of our sample for the Global South. For descriptive statistics, NAWE's rate is 7.33%. Although we could rely on UIS' average proportion for the several Global South regions in its database, we chose to calculate the mean presented in our sample due to the lack of data for emerging countries in UIS' database, which affects regions differently.

Loading our dataset into STATA v.14, we can calculate the mean proportion of the inbound mobility rate of the Global South countries in our sample. Using the 1.5 Interquartile Range (IQR) technique to remove outliers, we excluded seven[6] observations from a total of 77 countries. This left us with a mean statistic of 2.42% (similar to the world's 2.4%). We also found a sample standard deviation of approximately 2.28%. We then affirm that the mean proportion of Global South countries is less than NAWE's half. Even if we are dealing with proportions, since we calculated their average (descriptive statistics), we use the hypothesis test for the mean (μ) and not the proportion's π.

6 Except for Jordan, all the other six countries excluded are small Gulf or Caribbean states.

$$\mathbf{H_0}: \mu = 3.67\%$$
$$\mathbf{H_1}: \mu < 3.67\%$$

After calculating this one-tailed test at a 0.001 significance level,[7] the Z-score we got was approximately -4.14, way beyond the critical value to the left (-3.08). Therefore, we can reject the null-hypothesis and confirm that there is evidence to support the alternative one.

And just to strengthen this argument, we calculated another hypothesis test regarding the difference in the sample means. We contrasted the "Dataset_Emerging" mean with the one calculated from "Dataset_World" (more on those in the next section). The last one is comprised of 36 developed countries and, after applying the 1.5 Interquartile Range (IQR) technique, we were left with 34 observations, a sample mean of 8.73 and a sample standard deviation of approximately 5.03.[8]

$$\mathbf{H_0}: \mu_{Dataset\_Emerging} = \mu_{Dataset\_World}$$
$$\mathbf{H_1}: \mu_{Dataset\_Emerging} \neq \mu_{Dataset\_World}$$

We obtained a Z-score of -6.975, which indicates that it lies in the rejection zone to the left, way beyond any critical values at the 0.05, or 0.01 significance levels.

Descriptive Analysis and Data Presentation

We created two datasets for this study: "Dataset_Emerging" and "Dataset_World". The first comprises 77 observations from the Global South and the second includes those 77 plus 36 developed countries, totaling 113. In trying to identify the proportion of inbound students enrolled in a country's education system, we first used the descriptive statistics of the absolute share (that is, the number of international tertiary students divided by the total number of tertiary students in a given country), i.e., UIS' *inbound mobility rate*. However, we found that, due to the small number of tertiary students in some emerging countries, this would add bias toward less populated states.

For example, China, hosting almost 200 thousand international students, has a smaller rate (0.36%) than DR Congo (0.44%), which hosts only 2038 mobile

7 We used the formula $Z = \frac{\bar{X}-\mu}{\sqrt{\sigma^2/n}}$ to calculate the Z-score since, even though we do not know the population standard deviation, we are dealing with a large sample (n = 70).

8 The formula used was $Z = \frac{(\bar{X}_A-\bar{X}_B)-(\mu_A-\mu_B)}{\sqrt{\frac{s_A^2}{n_A}+\frac{s_B^2}{n_B}}}$.

students in the year analyzed. We ran one model using this rate as the dependent variable (Table 2), but the best way to capture a country's inbound mobility was to calculate the proportion of international students hosted regarding the total number of the sample's mobile students. Using this method, China gets its share of more than 10%, while DR Congo is at no more than 0.13%. We then assess the relevance of a country in the Global South scenario of international mobility, not its national proportion of foreign students. This technique also shields us against multicollinearity issues between the number of inbound students, the number of migrants, and the total population.

The explanatory variables for "Dataset_Emerging" are: *i.* spoken language; *ii.* relevance in university rankings; *iii.* migration network, *iv.* average distance; *v.* expected income; *vi.* hosting capacity; and *vii.* cost of living. Table 1 presents further details on them. For the first explanatory variable, we used a dummy in which '1' corresponded to a country that *de jure/de facto* speaks an official United Nations language (i.e. Arab, Chinese, English, French, Russian, or Spanish) or German (OECD 2011).[9] Secondly, academic rankings are common proxies for assessing higher education excellence in studies about inbound mobility. For this independent variable, we use ARWU/Shanghai ranking, but inclusively: the top 500 universities in 2017.

The third explanatory variable is proxied by the total number of international migrants in a country. Again, using the same logic applied for the dependent variable, we could take the national proportion regarding a country's population only, but it would create a strong bias toward less populated countries. We then divide the number of migrants in a given state by the total sum of migrants in all countries of our sample for the year 2015.[10] Average distance, the fourth explanatory variable, is proxied by the mean distance between a country's main economic center[11] (or its capital, which, for the majority of our sample, coincide)

9 As mobilities are generally performed in urban and academic environments, we applied an inclusive criterion. Some countries of our sample, such as former African colonies or Soviet Republics in Central Asia, speak a variety of languages. In situations in which the former colonial power language predominates in urban administrative, bureaucratic, and trade issues, we considered a positive result for the 'UN languages plus German' criterion. For example, India, Kazakhstan, Cote D'Ivoire, and Malaysia are countries in such a situation. All the ambiguous situations were analyzed using the CIA World Factbook and the CEPII (2011) dataset.

10 We do not use the values for 2017 since the UN has quinquennial publications and the year 2015 is the closest one with such data.

11 For example, we use São Paulo and not Brasília for Brazil, and Istanbul and not Ankara for Turkey. However, for the majority of the countries in the sample, the capital city and the main economic center coincide: Buenos Aires for Argentina, Santiago for Chile, Mexico City for Mexico, and so on.

and its ten closest destinations. As the literature points out, this relationship is negative and we selected the closest national entities and calculated the mean distance to check how far a country is from its neighbors. Expected income, the fifth variable, is measured by a country's GDP per capita in US Dollars in 2017. Hosting capacity, as the literature suggests, is proxied by the total population of a country in 2017. Lastly, the cost of living is proxied by World Bank's *Price level ratio of PPP conversion factor (GDP) to market exchange rate*, a continuous index number that considers the US equal to '1',[12] in 2017. "Dataset_World" has all those variables plus three dummies named i. *emerging*, ii. *col_pow45*, and iii. *oecd*. They were designed to capture country-specific factors that are present in Global South countries, based on a dummy created by Caruso and De Wit (2014) to differentiate Western and Eastern Europe.

We must highlight that we deal with pull aspects only and not dyadic factors (or, how country-A is attractive to country-B students given their distance, if they speak the same language, if A's expected income is higher, and so on), i.e., push-pull determinants. Therefore, we focus only on a country's attractiveness in general, not relative to some other country. We study these monadic factors mainly due to the lack of international mobility dyadic data for emerging states (for example, Mexico and Russia do not have the data on the nationality of the students hosted), which hinders dyadic analyses.

Furthermore, we chose the year 2017 for data collection since it is the most recent one (that predates the global COVID pandemic) with the highest availability of data for emerging countries in UIS' database. The indicators on international mobility are part of the *National Monitoring* series, which means that we are relying on a country's good-will to annually inform its numbers to Unesco. It not only hinders the creation of a panel or a time-series for the observations but forces us to include data from 2015 or 2016 as proxies for 2017 values for important countries in our sample (such as Israel and Egypt) that did not provide the 2017 indicators. For the developed countries, all the data on mobility is from 2017.

Lastly, regarding the inclusion criterion, "Dataset_Emerging" contains countries from Latin America, Africa, Asia (including the Middle East, the Caucasus, and Turkey), and European countries that are not members of the European Higher Education Area (EHEA), or even though are its members, were not part of the European Union (EU), the European Economic Area (EEA), or the European Free Trade Agreement (EFTA) in 2017 – that is, Belarus, Bosnia and Herzegovina, Moldova,

12 We did not use Numbeo's indexes due to their 'user input' methodology.

North Macedonia, Serbia, Russia, and Ukraine. "Dataset_World" comprises all the countries from "Dataset_Emerging" plus all the remaining countries from Europe, the United States, Canada, Japan, Australia, and New Zealand. All the observations were included if their inbound mobility data were available in UIS' database.

Table 1 summarizes and further explains the variables used in our models. The variables for distance (*dist10*), expected income (*income*), and hosting capacity (*population*) also have their data presented in natural logarithms (*ln_dist10*, *ln_income*, and *ln_population*). We performed this logarithmic transformation to obtain a more normalized dataset since these three variables are highly skewed.

Table 1 – Data Presentation

Variable	*Source*	*Definition*
inbound_rate	UIS (2020)	The proportion of foreign inbound tertiary students in a country divided by the total number of students in that country in 2017
inbound_sample	UIS (2020)	The number of foreign inbound tertiary students per country divided by the total sum of inbound students in our sample
Lang	CEPII (2011), CIA (2023)	Spoken language. '1' if a country's language is one of the official languages of the United Nations plus German, '0' if it is not
arwu500	ARWU (2017)	Number of universities in the top-500 ARWU/Shanghai ranking in 2017
mig_rate	UN DESA (2019)	International migrant stock (percentage of the total population) in 2015
mig_sample	UN DESA (2019)	Number of migrants per country divided by their total sum of our sample
dist10	CEPII (2011)	Mean distance between a country and the 10 closest national entities. This variable has a version with logarithmic transformation "*ln_dist10*"
income	UN (2017)	United Nations Stats 'GDP, Per Capita GDP – US Dollars' in 2017. This variable has a version with logarithmic transformation "*ln_income*"
population	UN (2017)	The total population of a country in 2017. This variable has a version with logarithmic transformation "*ln_population*"
liv_cost	World Bank (2020)	Price level ratio of PPP conversion factor (GDP) to market exchange rate in 2017. Index number (United States = 1)
emerging	De Wit and Caruso (2014)	The dummy takes the value of unity for countries in "Dataset_Emerging"
col_pow45	CEPII (2011)	The dummy takes the value of unity if a country had colonies or possessed territories that became independent in similar colonial relationship after 1945. It is based on the "col45" variable CEPII (2011)
oecd	OECD (2024)	The dummy takes the value of unity for OECD members in 2017

Source: Elaborated by the author (2024).

Main Findings

After data collection from several sources, the values were tabulated into two Microsoft Office Package 2019 Excel spreadsheets, each containing one dataset ("Dataset_Emerging" and "Dataset_World"). To run the OLS models, we used STATA v.14 software, and the commands and codes used are presented in a complementary script file submitted to the journal. The equation with the highest adjusted R-squared value – model (2) below – for "Dataset_ Emerging" can be described as follows. The variables are indexed by country "*i*" ($i = 1, [\ldots], 77$).

$$y inbound\_sample_i = \beta_0 + \beta_1 lang_i + \beta_2 arwu500_i + \beta_3 mig\_sample_i + \beta_4 ln\_dist10_i + \beta_5 ln\_population_i + \beta_6 ln\_income_i + \beta_7 liv\_cost_i + \varepsilon_i$$

Table 2 – Inbound Internationally Mobile Students in the Global South

	(1)	(2)	(3)	(4)
Dependent Variable	*inbound_rate*	*inbound_sample*		
lang	2.867524 (2.573091)	0.0026857 (0.0036851)	0.0021498 (0.0036834)	0.0019062 (0.0036369)
arwu500	0.1369286 (0.2524916)	0.0020745*** (0.0003656)	0.0019236*** (0.0003496)	0.0021348*** (0.0003631)
mig_rate	0.2309748** (0.094137)			
mig_sample		0.6558668*** (0.087333)	0.6717908*** (0.0869927)	0.7077024*** (0.0758839)
ln_dist10	-7.840511*** (2.56518)	-0.0007607 (0.0038138)	-0.000545 (0.0038315)	0.0000506 (0.0037632)
ln_income	0.3865359 (1.417735)	0.0024032 (0.0020221)	0.0009839 (0.001729)	
ln_population	-2.100384*** (0.7822091)	0.0013246 (0.0013134)	0.0015992 (0.0013044)	0.0015992 (0.0012403)
liv_cost	16.05242 (10.86713)	-0.02115 (0.015859)		-0.0112308 (0.0135247)
constant	79.97882*** (20.78983)	-0.0271534 (0.031483)	-0.0302044 (0.0315739)	-0.008717 (0.0274763)
Observations	77	77	77	77
Adjusted R-squared	0.4531	0.6973	0.6939	0.6955
F-Value	9.99	26.01	29.72	29.93

Note. Standard errors are presented in parenthesis. *significant at 10%. **significant at 5%. ***significant at 1%.

By analyzing the regressions' outputs, we can confirm that the model significantly improves when using *inbound_sample* as the dependent variable instead of *inbound_rate*. It is also interesting to note that hosting capacity (*ln_population*) has a negative coefficient in regression (1), which strengthens our argument that using the inbound rate proportion would create a bias toward less populated countries. Another problem with using *inbound_rate* as the dependent variable is found in *liv_cost*: it yields a positive coefficient, contrasting with previous works, probably due to the high inbound rate of the Gulf and Caribbean states, all of which present a high cost of living. Furthermore, from Table 1, we can infer that all statistically significant results in regressions (2), (3), and (4) are following the literature, that is, *arwu500* and *mig_sample*. In regression (1), *ln_dist10* and *mig_rate* present expected results, while *ln_population*, for the reasons stated, did not.

Moreover, to avoid common collinearity issues when measuring the cost of living, we conducted regressions (3) and (4) with either *ln_income* or *liv_cost*. The linear correlation between those variables is positive and moderately strong (Pearson's R = 0,61) and regression (4) yielded slightly higher adjusted R-squared and F-values than regression (3), but with a change in the coefficient for *ln_dist10*, contrasting with the literature. However, overall, even if presenting a lower F-value when compared to models (3) and (4), regression (2) produced the highest adjusted R-squared and all the (significant) coefficients in accordance with the literature.

The Global South Dimension

To improve our model and to reinforce our argument that emerging countries significantly differ from developed ones regarding their attractiveness, we increase the number of observations in the sample and add three new explanatory dummy variables. We now turn to the "Dataset_World". The new explanatory dummy variable named *emerging* takes the value of '1' when a country is also part of the "Dataset_Emerging". The *col_pow45* dummy takes the value of unity if a country had colonies or possessed territories that became independent after a similar colonial relationship since 1945, based on the "col45" variable present on CEPII (2011). Lastly, the third dummy, *oecd*, takes the value of unity if a country was an OECD member in the year 2017.

We designed these dummies based on Caruso and De Wit (2014) to try to capture a set of latent features that differentiate developed countries from emerging ones. For example, when choosing a Global South country as a destination instead of, say, Australia, several hard to measure variables may play a significant role, such as xenophobia, fear of violence, poor infrastructure, and even unrest about higher education excellence due to a feeble performance in rankings. This dummy tries to capture a socio-economic relationship in the Global South dimension.

Also, even if most of the emerging countries were colonies, a specific metric to evaluate the impact of recent/lasting colonial powers is interesting and has widespread application in the literature. However, based on CEPII (2011), we included not only the traditional and lasting colonial empires, such as the British and the French, but countries that lost territories that became independent countries since then (such as South Africa and Australia), the heirs of republic unions (such as the Russian Federation and Serbia), and countries that hold a dominion over some other territorial entity to this day (such is the case for Morocco). This dummy tries to capture a geopolitical relationship in the Global South dimension.

Lastly, due to the high availability of data and the institutional background linking Higher Education (more specifically international mobility) and OECD countries, distinguishing between this organization's members and other states seems a fruitful effort in testing for a Global South dummy. Sure, not all countries in the 'club of the rich' are developed, such are the cases of Chile, Mexico, and Turkey. However, due to a more 'institutional' approach fomented by the minds at the OECD and the *path dependence* of the organization, we opted to also include the emerging countries in OECD as positive cases for this dummy.

Table 3 displays the outputs of the regression models for "Dataset_World" with the inclusion of the three dummies. The equation with the highest adjusted R-squared and F-value – model (9) – for "Dataset_World" is described as follows. All the variables are indexed by "*i*" ($i = 1, [\ldots], 113$).

$$y inbound\_sample_i = \beta_0 + \beta_1 col\_pow45_i + \beta_2 lang_i + \beta_3 arwu500_i + \beta_4 mig\_sample_i + \beta_5 ln\_dist10_i + \beta_6 ln\_income_i + \beta_7 ln\_population_i + \varepsilon_i$$

Table 3 – Inbound Internationally Mobile Students in the World

Dependent Variable	(5)	(6)	(7)	(8)	(9)	(10)
	inbound_sample					
emerging		-0.0046366** (0.0022677)				
col_pow45			0.0071923*** (0.00195)		0.0071977*** (0.001924)	0.0073857*** (0.0019401)
oecd				0.0030197 (0.0024209)		
lang	0.0018903 (0.0014507)	0.0026073* (0.0014716)	0.0024168* (0.0013782)	0.0022028 (0.0014684)	0.0024192* (0.0013668)	0.0021949 (0.0013598)
arwu500	0.0007822*** (0.0001081)	0.0007548*** (0.0001073)	0.0007877*** (0.0001021)	0.0007779*** (0.0001079)	0.0007885*** (0.0000937)	0.0007701*** (0.0001006)
mig_sample	0.3881636*** (0.0643785)	0.3998374*** (0.0636816)	0.3662754*** (0.0611196)	0.3925821*** (0.0643063)	0.3658788*** (0.0578457)	0.3810935*** (0.0592626)
ln_dist10	0.0000755 (0.0013482)	0.0010699 (0.0014145)	0.0001037 (0.0012739)	0.0004147 (0.0013719)	0.0001064 (0.0012614)	-0.0000609 (0.001263)
ln_income	0.0010745 (0.0008314)	0.0005992 (0.0008515)	0.0007834 (0.0007896)	0.0008164 (0.0008547)	0.0007957 (0.000529)	
ln_population	0.0009951** (0.0004592)	0.0008778* (0.000456)	0.0005365 (0.0004513)	0.0007952 (0.0004852)	0.000535 (0.0004437)	0.0004838 (0.0004482)
liv_cost	0.0023953 (0.0049826)	-0.0010845 (0.0051954)	0.0000996 (0.0047489)	-0.0008528 (0.0056103)		0.0035837 (0.003197)
constant	-0.0267969** (0.0125074)	-0.0227745* (0.0124782)	-0.0170375 (0.0121107)	-0.0229148* (0.0128568)	-0.0170894 (0.0117991)	-0.0098933 (0.009737)
Observations	113	113	113	113	113	113
Adjusted R-squared	0.9024	0.9052	0.9128	0.9029	0.9136	0.9128
F-Value	148.88	134.74	147.61	131.45	170.31	168.58

Note. Standard errors are presented in parenthesis. *significant at 10%. **significant at 5%. ***significant at 1%.

The increase in the number of observations alone (5) improves the explanatory power of our model significantly. As expected, since all the independent variables were selected after a literature review of papers analyzing OECD countries. Furthermore, the inclusion of the three dummies also helped in increasing the models' adjusted R-squared (models 6,7,8,9, and 10) and F-values (models 9 and 10). Also, the *emerging* dummy has a negative impact on the dependent variable, while *col_pow45* and *oecd* are both positively related to *inbound_sample*.

Once again, we conducted two separate models (with either *ln_income* or *liv_cost*) to avoid collinearity issues. In this dataset, the linear correlation between the two variables is stronger (Pearson's R is equal to approximately 0.8). This time, the regressions with only one of the two variables (models 9 and 10) yielded higher adjusted R-squared and F-values than when the two were included (models 5, 6, 7, and 8), possibly due to this stronger linear relation. For the significant results, all of them produced coefficients similar to the literature (*lang*, *arwu500*, *mig_sample*, and *ln_population*).

An interesting finding is derived from *ln_dist10* outputs, even if not significant: the literature frequently presents a negative relationship, which makes sense. However, not using push-pull dyadic factors and just analyzing the attractiveness of a destination as a monadic factor may explain its positive coefficient in models (5,6,7,8, and 9): important inbound destinations are located in the 'edges' of the world, such as the US, Australia, Canada, Japan, Russia, Argentina, and China. As the literature often employs dyadic factors using OECD countries (which are mostly from Europe, a continent in which destinations are relatively closer when compared to the rest of the world), our result seems plausible when expanding the sample and focusing on pull factors only.

Also, observing our variables in Tables 2 and 3, one could speculate about the multicollinearity between the number of incoming students, the number of migrants, and the population of a country. Although the number of inbound international students may indeed contribute to an increase in the total number of migrants, which, in its turn, may aggregate in the total population, this does not happen in our sample. We dealt with this issue by using the variables *inbound_sample* and *mig_sample*, which do not compare the number of incoming students to the total number of students of a country, nor the proportion of migrants in a country's population, but their share in our sample. This eliminates expected multicollinearity effects when the number of inbound students affects the number of migrants and, successively, the population. Furthermore, we only used natural logarithm variables for distance, hosting capacity, and income in the models for the sake of standardization, and since the regressions, overall, yielded better outputs when doing so.

Lastly, out of the three dummies, the one that produced the best results was *col_pow45*. This is why we excluded income and cost of living (9 and 10) using this dummy. Not only it brought higher adjusted R-squared and F-values, it was significant at the 1% level – *emerging* was at the 5% level and *oecd* was not

significant in any of the measures. This strengthens the argument presented in the literature that a colonial link, or a recent domination relationship, boosts the flow of international students and that, perhaps, the geopolitical dummy is more suitable to distinguish the countries than the socio-economic ones (*emerging* and *oecd*). The results, however, were very similar and yielded comparable outputs.

Conclusion

Our conclusions can be summarized as follows:

i. The use of a larger dataset comprising the whole world presents better results than using emerging countries only. The inclusion of three dummies, the Global South dimension, also captures this.
ii. Most of the significant results in the ten models followed the literature. It is worth mentioning the importance of academic excellence (*arwu500*) and migration network (*mig_sample*) in all models – Caruso and De Wit (2014). Even though yielding a smaller number of significant results in the regressions, language (*lang*) and hosting capacity (*ln_population*) are also relevant to better understand inbound flows – Beine, Romain and Ragot (2014).
iii. The use of a given country's proportion of international students among its total number of enrollments (*inbound_rate*) proved to yield biased results when compared to the proportion of inbound students per country in the total number of mobile students in our sample (*inbound_sample*).
iv. Some variables, even though not significant, produced results that contrast with the current literature – Rumbley 2012; Didisse, Nguyen-Huu and Tran 2018. That is the case for *ln_dist10* and *liv_cost* (Table 3). The first case can be explained by the monadic approach – instead of a dyadic one – and the inclusion of non-European countries. The second one is harder to grasp, but due to its high collinearity with *ln_income*, when excluding the cost-of-living variable from "Dataset_World" – model (9) –, the output displayed better results.

This article tried to address a lack in the literature and on databases on emerging countries' inbound mobility. Using *inbound_sample* to measure the

dependent variable, adopting more inclusive criteria when estimating the pull factors, and creating the *emerging*, *col_pow45*, and *oecd* dummies proved to be effective techniques not only to recognize the need for a different approach to Global South countries but also to design a quantitative model including them.

The lack of data regarding international mobility in emerging countries is blatant and we tried to remedy this by choosing the year in UIS' database in which data for the Global South is more abundant, which forced us to create a cross-section and use OLS techniques. Sure, this can make us wonder if our results were due to the sheer luck of choosing an atypical year for the international mobility scenario.

However, three reasons can advocate for the relevance of the 2017 results when compared to other studies: *i.* the (quasi-)constant nature of the share of international students per country, their languages, population, and distance (Didisse, Nguyen-Huu and Tran 2018); *ii.* the reduced number of observations (77 and 113) in our sample – other studies with such a limitation, such as Kahanec and Králiková (2011) and Caruso and De Wit (2014), use OLS as well; and *iii.* our results were in accordance with the literature that uses panels and/or time-series for modeling OECD/European countries, who benefit from more abundant data. However, if more (dyadic) data for emerging countries become available in the next years, the creation of a panel or a time-series may help us better understand the evolution of the pull factors presented here. Also, future research that overcomes the push-pull dynamics, which are declining in migration studies, are welcome.

Overall, our models yielded a satisfactory number of statistically significant explanatory variables and good adjusted R-squared and F-values and the dummy that seems to better capture the Global South dimension is *col_pow45*. These results can and should help policymakers when designing national policies to attract inbound students and boost HEI internationalization in emerging countries.

References

Beck, Kumari; Michelle Pidgeon. 2020. Across the Divide: Critical Conversations on Decolonization, Indigenization, and Internationalization. In International Education as Public Policy in Canada, 384-406. Edited by Merli Tamtik, Roopa Desai Trilokekar, and Glen A. Jones. Québec, Canada: McGill Queen's University Press.

Beine, Michel; Romain, Noël; Ragot, Lionel. 2014. Determinants of the international mobility of students. Economics of Education Review, no. 41: 40–54.

Buckner, Elizabeth; Sharon Stein. 2019. What Counts as Internationalization? Deconstructing the Internationalization Imperative. Journal of Studies in International Education 24, no. 2: 1-16.

Caruso, Raul; De Wit, Hans. 2014. Determinants of Mobility of Students in Europe: Empirical Evidence for the Period 1998-2009. Journal of Studies in International Education 19, no. 3: 1-18.

CEPII. 2011. CEPII's distances measures: the GeoDist Database CEPII. Paris, France.

Chen, Tse–Mei; Barnett, George A. 2000. Research on International Student Flows from a Macro Perspective: A Network Analysis of 1985, 1989 and 1995. Higher Education 39, no. 4: 435-453.

CIA. 2023. The World Factbook, Field Listing: Languages. https://www.cia.gov/library/publications/the-world-factbook/fields/402.html Retrieved October 28, 2020.

Didisse, Jonas; Nguyen-Huu, Thanh T.; Tran; Thi A-D. 2018. The Long Walk to Knowledge: On the Determinants of Higher Education Mobility to Europe. The Journal of Development Studies 55, no. 6: 1-21.

Docquier, Frédéric; Rapoport, Hilel. 2012. Globalization, brain drain and development. Journal of Economic Literature 50, no. 3: 681–730.

Junqueira, Cairo G. B.; Baldrighi, Rafael de M. 2020. A internacionalização do ensino superior no Brasil analisando comparativamente a mobilidade internacional de estudantes face à realidade latino-americana. Carta Internacional 15, no. 3: 29-54.

Kahanec, Martin; Králiková, Renatá. 2011. Pulls of international student mobility. IZA Discussion Paper No. 6233, Bonn, Germany: IZA.

Mazzarol, Tim; Soutar, Geoffrey N. 2000. Push-pull Factors Influencing International Students Destination Choice. The International Journal of Educational Management 16, no. 2: 82-90.

Mignolo, Walter D. 2002. Geopolitics of Knowledge and the Colonial Difference. The South Atlantic Quarterly 101, no. 1 57-96.

OECD. 2011. Education at a glance 2011, Paris, France.

OECD. 2019. Education at a glance 2019, Paris, France.

Pedro, Eugénia; Franco, Mário. 2015. The importance of networks in the transnational mobility of higher education students: Attraction and satisfaction of foreign mobility students at a public university. Studies in Higher Education 41, no. 9: 1627–1655.

Rosenzweig, Mark R. 2008. Higher education and international migration in Asia: Brain circulation. Annual World Bank conference on development economics, 59–100.

Rumbley, Laura. 2012. So many data, so little clarity. In Internationalisation Revisited: New dimensions in the Internationalisation of Higher Education, 125-134. Edited by J. Beelen and H. De Wit. Amsterdam, The Netherlands: CAREM.

Un Desa. 2019. United Nations Population Division. https://www.un.org/en/development/desa/population/migration/data/estimates2/countryprofiles.asp Retrieved October 16, 2020.

Unesco. 2020. Unesco Institute For Statistics. http://data.uis.unesco.org/ Retrieved October 20, 2020.

Van Bouwel, Linda; Veugelers, Reinhilde. 2013. The determinants of student mobility in Europe: The quality dimension. European Journal of Higher Education 3, no. 2: 172–190.

World Bank. 2020. Price level ratio of PPP conversion factor (GDP) to market exchange rate. https://data.worldbank.org/indicator/PA.NUS.PPPC.RF?end=2019&start=2013 Retrieved November 5, 2020.