Who offers English-medium instruction? Exploring university characteristics with random forests

(article starts on next page)

Who offers English-medium instruction? Exploring university characteristics with random forests

Peter Wingrove

View supplementary material

Published online: 05 Oct 2025.

Submit your article to this journal

Article views: 746

View related articles

View Crossmark data

Routledge
Taylor & Francis Group

Who offers English-medium instruction? Exploring university characteristics with random forests

Peter Wingrove

Department of Communication and Learning in Science, Chalmers University of Technology, Gothenburg, Sweden

ABSTRACT

English-medium instruction (EMI) has spread throughout European higher education since the turn of the century. Whilst much valuable research has been conducted on pedagogic concerns, the growth of EMI across countries, and macro-level drivers, we know surprisingly little about the kinds of institutions that offer EMI. This paper fills this gap by exploring the meso-level (institutional) predictors of EMI in European higher education institutions using interpretative machine learning. Random forests are employed to analyse data from the European Tertiary Education Register and Study Portals, focusing on meso-level features linked to EMI adoption. The model achieves an accuracy of 84% and a Cohen's Kappa of 0.64, indicating the strong predictive performance of meso-level features. The most important feature was PhD discipline diversity, suggesting that discipline-diverse institutions were more likely to offer EMI than specialist institutions. An exploration of all features suggests that the archetypal EMI-offering institution in Europe is disciplinarily diverse, large, internationally oriented, research intensive, resource rich, and postgraduate focused. These findings are discussed through the lens of institutional isomorphism and the global spread of EMI, with divergence in the case of EMI potentially giving rise to elitism or openness, and the implications this has for higher education planning and policy.

Introduction

English-medium instruction (EMI) is a runaway train, but who is driving? Since the turn of the century EMI has expanded rapidly in European higher education, with English-taught degree programmes (ETPs) tripling from 8,000 in 2013 to 24,000 in 2023/24 (Wingrove et al. 2025). Once concentrated in central and northern Europe, EMI had become a pan-European practice by 2023/24, with around half of all higher education institutions (HEIs) offering EMI, up from a quarter a decade prior (Wingrove et al. 2025). EMI is commonly defined as the use of English to teach academic subjects in non-English L1 (first language) countries (Macaro 2018). Whilst this definition is much cited, it does not tell us about the ontological nature of EMI: its nature as a fluid set of educational, linguistic, and institutional practices that shape and are shaped by the contexts in which they occur.

As a starting point, EMI is an instructional setting (Pecorari and Malmström 2025) not a pedagogic model. This makes the spread of EMI intriguing as whilst we have examples of pedagogic models spreading throughout higher education such as peer learning (Stigmar 2016), the spread of an instructional setting appears to be a new phenomenon. This suggests a different set of drivers,

---

i.e. not driven by teachers through pedagogic motivations – indeed, there are cases where faculty have *resisted* EMI adoption (Murphy and Zuaro 2021) – but rather by other forces.

Prior research has therefore pointed to macro-level political and economic drivers to explain EMI uptake. This includes marketisation (Macaro 2018), internationalisation and the Bologna Process (Wilkinson and Gabriëls 2021), massification (Yuksel et al. 2024), the global knowledge economy (Hultgren 2024), and neoliberal principles of governance (Hultgren et al. 2023; Wingrove et al. 2024). While such macro-level forces illuminate the diffusion mechanisms underpinning EMI's rapid cross-country expansion, they represent only one side of the puzzle; equally important are the meso-level dynamics, where institutions differentially interpret, adopt, resist, or even reinforce these pressures, shaping the trajectory of higher education's linguistic landscape in distinct ways.

This is especially important given the ethical concerns raised about EMI. To name a few: EMI is not driven by pedagogic concerns and therefore education quality may be sacrificed (Macaro et al. 2018); EMI can be traumatic for participants coerced into conducting their studies and careers in an L2 (second language) (Block 2022); and EMI can be threatening to local cultures and languages (Murphy and Zuaro 2021; Wilkinson 2013). A key concern for both the higher education landscape and the findings of this study is that EMI is increasingly recognised as an elite practice, associated with prestige, international rankings, and multinational business networks (Kuteeva and Kaufhold 2024). This creates divisions between participants, disciplines, and institutions, and erects linguistic barriers to entry for L2 speakers, furthering elite stratification in higher education. However, an opposing view sees a *lingua franca* as building bridges and enabling access to higher education to a wider range of students.

Importantly, these concerns and debates highlight that EMI is not merely a neutral policy response to global trends, but a practice that carries linguistic, psychological, educational, and societal consequences. Understanding these consequences requires moving beyond macro-level accounts to examine institutional patterns of adoption and, in particular, to identify where divisions emerge across different types of institutions. Such an analysis is central to assessing how EMI reshapes the higher education landscape, from patterns of access and participation to the distribution of resources and opportunities across institutions. It also offers a means of refining theoretical accounts of EMI expansion, which have thus far been dominated by system-level drivers.

Therefore, this study investigates the spread of EMI at the institution level, identifying which institutional characteristics are associated with the adoption of EMI. Two comprehensive datasets underpin the analysis: the European Tertiary Education Register (ETER) (Wagner-Schuster et al. 2023), which provides extensive data on revenue, expenditure, students, personnel, research activity, Erasmus, indicators (e.g. Herfindahl index: a measure of subject-specialisation), and classifications (e.g. institutional control, size); and Study Portals, which offers data on internationally advertised EMI programmes (Wingrove et al. 2025). These combined datasets enable an exploration of the relationship between HEI characteristics and EMI adoption.

Methodologically, this study employs random forests, a non-parametric machine-learning algorithm that excels at identifying patterns in large datasets (Breiman 2001a). Random forests construct ensembles of decision trees, each trained on a random subset of the data and considering a random selection of features at each split, to classify or predict outcomes. By aggregating results, random forests improve accuracy and reduce overfitting. This method is particularly well-suited for identifying the most influential variables in a dataset, as it ranks features by their importance based on their contribution to reducing uncertainty across the forest. The 'forest' here is the collection of decision trees built on random subsets of data and features. For researchers working with complex institutional data, such as those in this study, random forests provide a powerful and interpretable tool for uncovering meaningful patterns and insights.

Whilst random forests excel at predictive performance, in this study an interpretative machine learning perspective is adopted (Molnar 2022), focusing not only on model performance but also on uncovering the institutional features linked to EMI adoption. When all features are considered together, we can build profiles for the archetypal institutions that offer and do not offer EMI. This

aligns with emerging research practices in social science that emphasise the integration of machine learning techniques for concept discovery and prediction (Grimmer, Roberts, and Stewart 2021). Consequently, this paper addresses the following research questions:

1. In a random forests model, how effective are institutional characteristics for predicting the presence or absence of EMI in higher education?
2. Which institutional characteristics are most important for predicting the presence or absence of EMI, and what relationships do they exhibit with EMI adoption?
3. Given these findings, what types of institutions are most likely to offer and not offer EMI?

Our first research question concerns model performance and would give us an indication of the relevance of institutional characteristics for predicting EMI. Model performance is assessed by both accuracy (% correct cases) and Cohen's Kappa.

The second question explores model interpretation to identify how key meso-level features, related to students, faculty, finances, and more, shape EMI adoption. Gini scores are used to rank feature importance, partial dependence plots to show how individual features affect predictions, and breakdown plots to illustrate specific cases.

Our final research question infers which institutions offer EMI based on model performance, feature importance, and feature–EMI relationships. One advantage of the current method is that institutions do not necessarily fall into distinct categories, but rather blend characteristics, which our method excels at identifying. This provides a highly detailed, robust, and granular exploration of institutional features that predict EMI, ranked in terms of importance, rather than simplistic descriptive statistics, which could be misleading due to underlying correlations among institutional characteristics.

This study discusses findings through the lens of institutional isomorphism and considers their implications for framing EMI as either elite or inclusive. Recommendations are offered for policymakers and institutional leaders.

Materials and methods

Due to the large amount of data in the study, the supplementary materials contain full details on catchment (Appendix A) and features (Appendix B). This includes descriptive statistics, missingness, and details on feature engineering per feature. Note that Appendix B contains the technical names for features to ease comparability between studies and the study employs a more intuitive naming convention for readability.

Data

This study utilises two robust data sources to model the predictors of English-taught degree programmes (ETPs) in European higher education institutions (HEIs). The first source is the European Tertiary Education Register (ETER) (Wagner-Schuster et al. 2023), a comprehensive dataset of approximately 3,500 HEIs in the European Higher Education Area that comprises the predictors. ETER provides data on: institutional descriptors, geographic information, expenditures, revenues, personnel, students, graduates, research, indicators, classifications, Erasmus data, EQAR (European Quality Assurance Register), and EU-FP (European-Union Framework Programmes) project data. ETER has been employed in studies and policy briefs exploring European higher education (e.g. Painsi, Leitner, and Rybnicek 2024). ETER provides data from 2011–2021. The 2019 dataset was selected to prioritise modern data, to avoid the pandemic disruption from 2020 onwards, and to align with our Study Portals data described in the following paragraph. In other words, 2019 represents the latest year of 'normal' higher education operations available to the researchers before COVID-19 caused disruption to mobility, enrolment, and programme advertisements.

The second source is Study Portals, which offers data on the outcome variable: English-taught degree programmes (ETPs). Study Portals is the largest online database of ETPs worldwide, where universities advertise their degree programmes to prospective students. All programmes listed on Study Portals are delivered in English. One strength of Study Portals is that it provides a standardised measure for ETPs across the EHEA. Combined, these datasets provide a solid foundation for analysing the institutional characteristics associated with EMI.

Dimensionality reduction was applied to the 700 ETER features to mitigate multicollinearity, enhance interpretability, and stabilise feature importance in the random forest model. While random forests tolerate high-dimensional data, removing redundant or highly correlated features improves efficiency and highlights meaningful predictors. As this is a theory-building study, features were retained unless there was a clear reason for removal. Specifically, (1) descriptors and geographic data (e.g. institution name, location) were removed and replaced with ID numbers; (2) flags and notations containing purely textual information were excluded; (3) continuous variables were retained over equivalent categorical versions; (4) categories with Variance Inflation Factors >10 were excluded, including all 'graduates' features, which were collinear with enrolment data; and (5) variables with >60% missingness were removed, balancing information loss against the need for retention in large-scale, theory-building analyses (Allison 2001; Madley-Dowd et al. 2019). This process reduced the dataset to 49 features across nine categories (expenditure, revenue, personnel, students, research, indicators, classifications, Erasmus, and EU-FP projects). ETER handbook names are retained in the appendix for comparability, except for engineered features, which mandate slight name alterations.

Prior to running analyses, feature engineering was conducted to improve model performance and interpretability. ETP offerings were modelled as a binary outcome. To address sampling zeros – years where institutions did not advertise ETPs despite doing so in adjacent years – institutions were classified as EMI providers if they advertised programmes within a one-year window on either side of the academic year 2019/20. This approach aligns with the study's aim of predicting whether institutions provide EMI, rather than quantifying programme numbers, which can be misleading due to variability in programme enrolment and length. Strictly speaking, this models whether an institution *sometimes* or *always* advertises EMI against those that *never* advertise EMI. Further details on feature engineering are provided in Appendix B, covering log-transformations and standardisations.

Analysis

To model the presence of English-taught programmes (ETPs) at European higher education institutions, a random forests classifier was employed (Breiman 2001a, 2001b). The analysis was implemented in R, using the tidymodels framework (Kuhn and Silge 2022). Given the binary outcome variable (ETP presence vs. absence), the model was configured for classification.

Before training the model, standard preprocessing steps were applied to ensure data consistency. All numeric predictors were normalised, and missing values were imputed using k-nearest neighbours. Prior studies have suggested the value of imputation with this dataset (Bruni, Daraio, and Aureli 2021). The dataset was split into 80% training and 20% test data, stratified by the outcome variable to maintain proportional class distribution. The training set was used to train the model, which was subsequently applied to the test set for final model performance and interpretation.

The Random Forest model was constructed using the ranger engine (Wright and Ziegler 2017), optimised for computational efficiency. It was set to include 500 decision trees, with hyperparameter tuning performed on the number of randomly selected features and the minimum number of observations in a terminal node. In this study's model, the minimum number of trees was 19 and the minimum number of observations in the terminal node was 9. A latin hypercube sampling strategy was used to generate 100 candidate hyperparameter combinations. The optimal hyperparameter values were selected based on a 10-fold cross-validation with 3 repeats conducted on the training data.

The model performance was assessed on the test set using the classification metrics 'accuracy' and 'Cohen's kappa'. Both accuracy and kappa were used as the dependent variable has imbalance: 63.2% (class 0: non-EMI) versus 36.8% (class 1: EMI), which means that accuracy alone may not fully reflect model performance. In this context, kappa provides an additional evaluation metric, accounting for agreement occurring by chance. For context, kappa score ranges are: slight agreement (0–0.2), fair agreement (0.21–0.4), moderate agreement (0.41–0.6), substantial agreement (0.61–0.8), and almost perfect alignment (0.81–1) (Landis and Koch 1977).

Precision and *recall* were also included to further assess the model's ability to correctly identify instances of both classes. Precision measures the proportion of true positive predictions among all positive predictions, while recall evaluates the proportion of actual positives that were correctly identified by the model. Together, these metrics provide a more comprehensive view of model performance, ensuring that both false positives and false negatives are considered when evaluating the effectiveness of the classification model.

Feature importance was analysed using Gini importance scores derived from the random forest model. Gini importance measures the contribution of each variable to the model's predictive accuracy by quantifying the average decrease in node impurity when the variable is used for splitting. Higher scores indicate features that play a more significant role in distinguishing between HEIs that do and do not offer EMI.

Following this, features were explored using partial dependence plots (PDPs). PDPs illustrate the marginal effect of a given feature on the predicted probability of an outcome, holding all other variables constant. This approach allows us to isolate and visualise how changes in individual predictors influence the likelihood of a HEI offering EMI.

Finally, breakdown plots are used to show how predictions for individual cases are influenced by each feature. Breakdown plots decompose the prediction for each observation into contributions from each feature, offering a transparent view of how the model arrives at its decision for specific cases. This technique is useful for interpreting the impact of each predictor in a clear, case-specific context. By combining these methods, we can better understand the key meso-level features driving the probability of HEIs offering EMI.

Findings

The findings presented here first cover model performance, which tells us how effective institutional characteristics are for predicting the presence or absence of EMI (RQ1). Second, we turn to feature interpretation which is comprised of feature importance and an exploration of feature relationships with EMI outcomes. This tells us which characteristics are most important in our model and what relationships they exhibit with EMI adoption (RQ2). This process is then illustrated through break down plots concerning individual cases. Considering all of these findings, we are able to paint a picture of the archetypical institutions which offer and do no offer EMI (RQ3).

Model performance

Beginning with our RQ1 on model performance. The confusion matrix for model performance can be seen in Table 1. The random forest model demonstrated a high level of accuracy in predicting

Table 1. Confusion matrix: predictive accuracy of the random forest model for EMI.

	Actually non-EMI	Actually EMI	
Predicted non-EMI	321 [true negatives]	60 [false negatives]	
Predicted EMI	32 [false positives]	146 [true positives]	Precision: 82%
		Recall: 71%	**Accuracy:** 84% **Kappa:** 0.64

whether European higher education institutions offered English-medium instruction (EMI), achieving an overall **accuracy of 84%** and a **Cohen's Kappa of κ = 0.64** on the test set. This means that the model was correct in 84% of cases, and Cohen's Kappa indicates 'substantial' agreement. Both these measures show strong predictive performance, answering the first research question: institutional characteristics are relevant and strong predictors of EMI.

Further, the precision score shows that when the model predicts that an institution offers EMI, it is correct 82% of the time. Recall (also known as 'sensitivity') is lower at 71%, which means that the model incorrectly classifies 29% of EMI institutions as non-EMI. This means that the model favours precision over recall, reducing false positives (type 1 errors) at the cost of increasing false negatives (type 2 errors).

Model interpretation

Feature importance

Now we turn to RQ2, on feature importance and an exploration of feature relationships with EMI outcomes. Feature importance is first assessed using Gini scores from the random forest model (Figure 1). In answer to RQ2, the top ten predictors were: PhD discipline diversity, the number of EU funded projects, the number of academic staff, the share of foreign students, the total number of students, the share of doctoral students, total expenditure, PhD orientation, total revenue, and the share of students in business, administration, and law. This shows a mix of features related to discipline diversity, internationalisation, finances, and the doctoral level.

Several features one might assume to be relevant, such as institutional control (public or private ownership) and education intensity, were shown to have low importance, as did most disciplinary categories. While Gini scores reveal importance, they do not convey directionality, which are explored via PDPs in the next section.

Feature exploration: partial dependence plots

Figures 2–4 show Partial Dependence Plots (PDPs) for the top 30 features in our model in rank order. PDPs illustrate the marginal effect of a given feature on the predicted probability of an outcome. This approach allows us to isolate and visualise how changes in individual predictors influence the likelihood of a HEI offering EMI. Due to space constraints, the full set of plots are given the appendix, which includes ranks 31–49. In this section, the top 30 PDPs are unpacked thematically.

Discipline diversity is a key theme in our data. 'PhD discipline diversity' and 'discipline diversity' for BA and MA students appear at rank 1 and rank 19, respectively, both linking discipline diversity with EMI and discipline specialisation with non-EMI. Taking both features together and considering discipline diversity across the entire university, why is this so important? A face-value interpretation is that institutions with a wider range of disciplines simply have more points of entry for EMI to take hold.

Another interpretation is that disciplinary diversity signals something significant about institutional identity and agenda. Discipline-diverse institutions tend to adopt a 'big tent' approach to higher education, shaped by massification: bringing together large numbers of students from a range of disciplinary, social, and national backgrounds. This inclusivity requires not only curricular breadth but also communication strategies that extend across internal and external boundaries. English plays a pivotal role in this regard, serving as the *lingua franca* through which institutions position themselves as outward-facing and international, while also facilitating cross-disciplinary interaction within the campus. The use of English here is not incidental but central to the institutional agenda: it supports the aim of attracting a broad student body and of projecting an international identity.

Conversely, subject-specialised institutions, such as medical colleges or law schools, have a different agenda. They aim to train students to enter a specific national-level workforce, with programmes tailored to the standards, practices, and language of their home country. Indeed, the

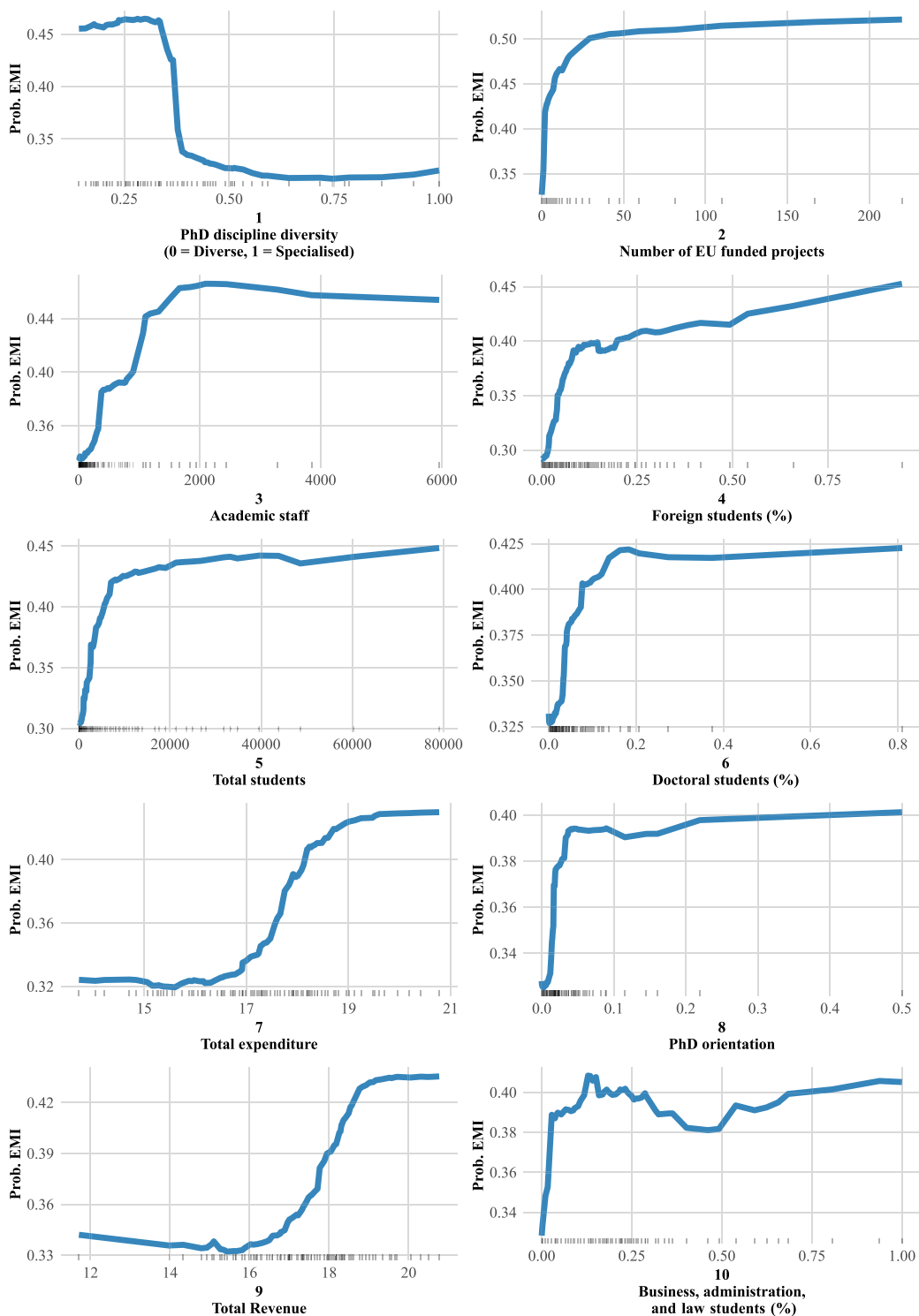**Figure 1.** Feature importance plot for English-medium instruction.

Figure 2. Partial dependence plots for top 10 features.
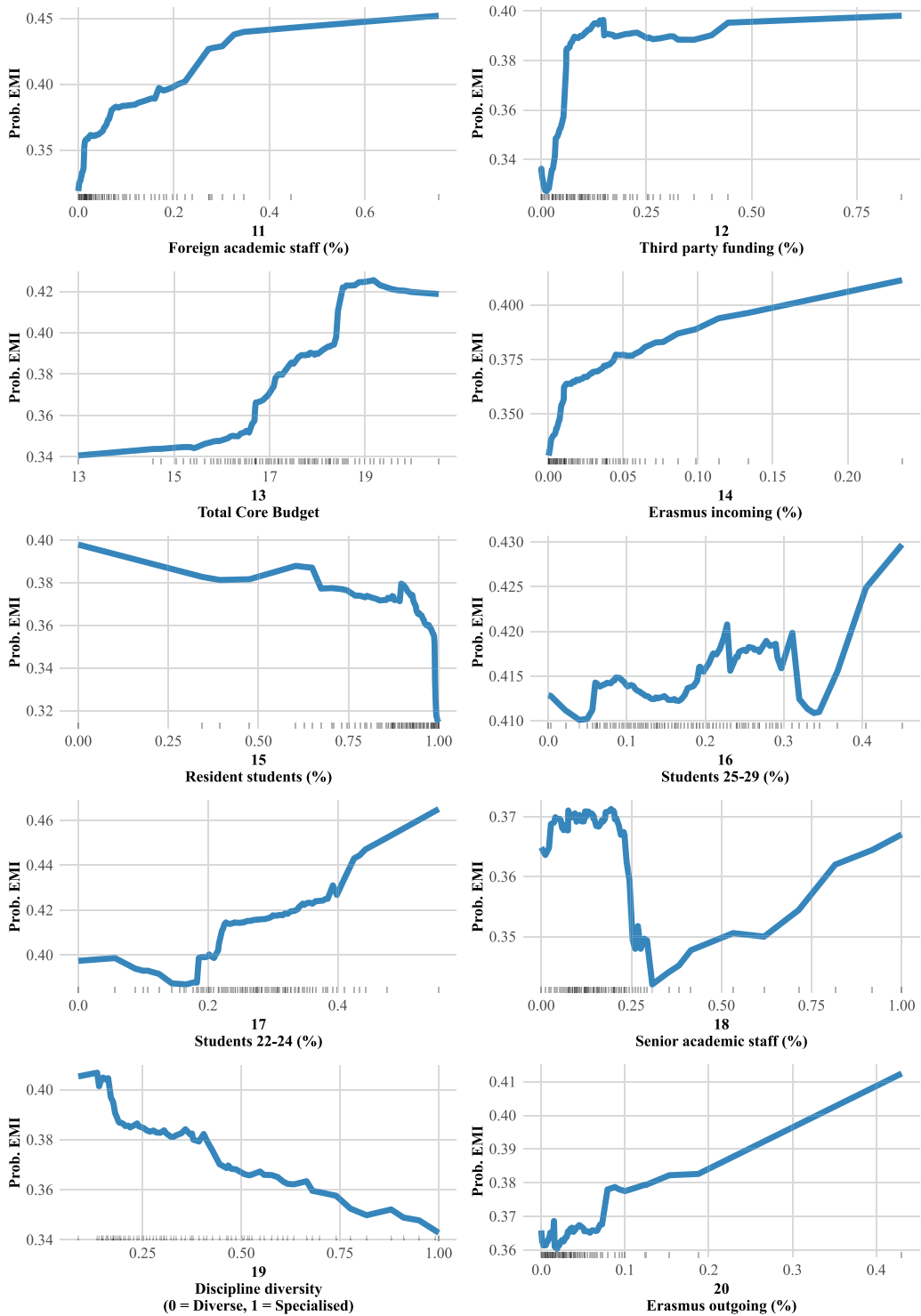
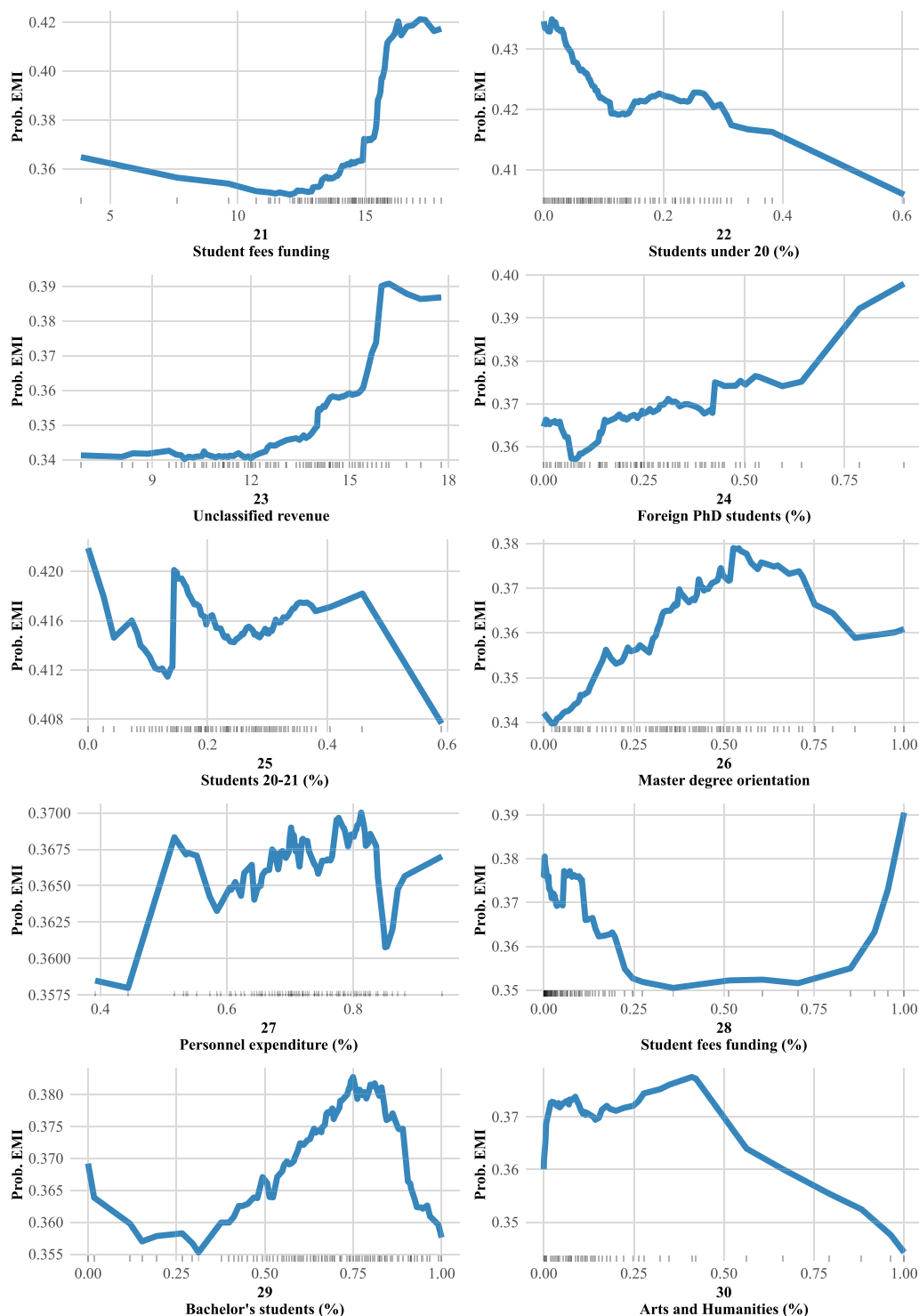**Figure 3.** Partial dependence plots for rank 11–20 features.

Figure 4. Partial dependence plots for rank 21–30 features.

nationally defined accreditation standards of legal and medical education often render such degrees applicable only within the host country's jurisdiction, rendering such degrees undesirable to those who wish to study abroad and return home.

Institution size is an important factor in our model. This is captured by two features: total academic staff (rank 3) and total students (rank 5), with larger institutions being more likely to offer EMI. Similar to our interpretation of discipline diversity, a face-value interpretation suggests that larger institutions will offer more programmes and this increases the likelihood that at least one will be English-medium. However, we can also consider the fact that EMI itself is used as a strategic device to increase enrolment, such as in the case of the Netherlands, where EMI was used to meet enrolment targets once the Dutch market had been saturated (Hultgren and Wilkinson 2022). Therefore, EMI may align with institutional goals concerning massification and is therefore more likely to be found in larger institutions.

Our model shows that EMI is associated with research intensive institutions. Within the top ten features, three are associated with the doctoral level: PhD discipline diversity (rank 1), the share of doctoral students (rank 6), and PhD orientation (rank 8), which is an indicator based on graduation rates. This contrasts with education intensity (rank 45), a student-staff ratio, which appears unimportant. This is surprising as EMI is a pedagogic phenomenon and our measure of EMI concerns bachelor's and master's programmes. However, it appears that research orientation is much more important than pedagogic focus when it comes to predicting the presence of EMI within an institution. It may be the case that research measures are a proxy for international cooperation on research projects and competition for international talent, factors closely linked to EMI adoption. Moreover, doctoral students face pressures to develop an international profile, necessitating advanced English proficiency, which in turn lowers barriers to EMI teaching and increases their likelihood of entering teaching-related roles, further explaining the close link between the doctoral-level and EMI provision.

EMI is predicted by our financial features. This includes revenue (rank 7), expenditure (rank 9), total budget (rank 13), student fees funding (rank 21), unclassified revenue (rank 23), and share of personnel expenditure (rank 27). Except for personnel expenditure, in all cases we see clear positive relationships, with greater resourced institutions offering EMI. Some of these relationships are intuitive: for example, the link to student fees funding aligns with theories on the marketisation of higher education. Institutions appear to be treating EMI as a market-responsive product designed to attract fee-paying international students. Moreover, the fact that there is no clear positive or negative relationship concerning share of personnel expenditure challenges assumptions that EMI significantly strains personnel budgets, more likely, this is counter-balanced by revenue generated by EMI. Consequently, the capacity to generate revenue through fees provides both the financial incentive (motivator) and the necessary capital (facilitator) to develop and sustain EMI programmes. One surprising finding is that EMI is also predicted by increased expenditure, again suggesting that EMI is associated with universities of a certain type: large and well resourced.

Internationalisation is also key to the provision of EMI. We see this with EU funded projects (rank 2), the share of foreign students (rank 4), the share of foreign academic staff (rank 11), Erasmus incoming (rank 14), Erasmus outgoing (rank 20), and foreign PhD students (rank 24). In all cases greater internationalisation is associated with EMI. Whilst this is unsurprising, there are a few novel contributions to note. First, internationalisation does not scale perfectly linearly. Instead we see inflection points for our predictors. In other words, internationalisation is only a strong predictor up to point: in the case of foreign students the inflection point occurs at approximately 10–12%. This threshold could be explained as identifying the expected share of international students required or attracted by EMI (10–12%). Alternatively, these thresholds may identify changes in institutional strategies, aiming for certain levels of internationalisation which EMI attracts. However, since our model is binary these should be seen as indicative rather than definitive cut-offs; modelling EMI as a count could reveal more continuous scaling and weaker threshold effects. Second, it is interesting how the most important measure of internationalisation, the number of EU funded projects, also acts as an indicator of research and finances, again emphasising the importance of these two factors.

Postgraduate orientation, towards the master's level (rank 26) and the doctoral level (rank 6), is associated with EMI. In contrast, an exclusive focus on the bachelor's level (rank 29) lowers the likelihood of offering EMI. This aligns with current figures which show that EMI is typically a master's level phenomenon in Europe (Wingrove et al. 2025). It is not surprising then that a focus on the master's level is linked to a higher probability of EMI. What is surprising, however, is that doctoral enrolment is a more important feature.

The disciplinary categories in our model occur at rank 30 or below, with the exception of 'business, administration, and law'. This is likely due to the inclusion of 'business', which is the most popular EMI discipline (Wingrove et al. 2025). There is a data limitation here with business and law grouped within a single feature, which almost certainly differ in terms of EMI representation. However, this does provide a general picture of which kinds of institutions offer EMI. For instance, over specialisation in 'health and welfare', 'arts and humanities', 'education', and 'engineering, manufacturing, and construction' lowers the probability of an institution offering EMI. The same is not true for 'business, administration, and law', 'ICT', and 'social sciences, journalism, and information' (see Appendix C). This echoes some of the discussion concerning discipline diversity versus specificity: certain kinds of institutions will specialise and focus on local markets and industries, whereas others will look outwards and use English as a bridge towards internationalisation. This orientation is not independent from discipline, which may in fact be a crucial component in deciding the extent and nature of internationalisation agendas.

Several lower ranked features provide some surprising insights. In our model EMI is predicted by more part-time students (rank 39), gender balance (ranks 31, 32, and 36), older students (ranks 16 and 17), and older institutions (rank 34). A greater share of older and part-time students aligns with the fact that EMI is primarily a postgraduate phenomenon, with older students already in work deciding to study part-time. It is possible that universities with a balanced gender profile are also those that actively recruit more broadly across demographics, including internationally, which supports the adoption of EMI. A balanced gender profile, therefore, may be an indicator of a broader institutional commitment to diversity and inclusive recruitment strategies, which inherently aligns with the global and accessible nature of EMI programmes.

Breakdown plots: predictions for two cases

In this section, two breakdown plots are explored to demonstrate how the model generates predictions for EMI (Figure 5) and for non-EMI (Figure 6). In Figure 5, total academic staff, total students, total revenue, total expenditure, the number of EU funded projects, and the share of foreign academic staff all paint a picture of a large well-resourced international university, which increases the probability of the institution offering EMI, according to our model. Several factors diminish this probability: scarce participation in Erasmus, a gender imbalance at the PhD level, and a low share of student fees funding as a proportion of total revenue. Overall, the model assigns a probability of 0.596 for offering EMI and allocates this as a case of EMI.

In Figure 6 the model predicts that the institution does not offer EMI. This is due to disciplinary specialisation at the PhD level, the absence of students in 'business, administration, and law', and 'ICT', no participation in the Erasmus programme, and a low number of students and academic staff. Several factors increase the likelihood of offering EMI, such as foreign academic staff, a high share of third party funding, and a generally older student body, however, this is not enough to outweigh the other factors that suggest non-EMI. Although grouped under 'all other factors', so not visualised in the break down plot, the share of students in engineering, manufacturing, and construction is 88.56% for this case, demonstrating specialisation in this area. This appears to be a small technical college geared towards national level industry, which would be less motivated to offer EMI.

The archetypal European English-medium instruction university

In answer to RQ3, our archetypical European EMI university is: disciplinarily diverse, large, internationally oriented, research intensive, resource rich, and postgraduate focused. The opposite is true
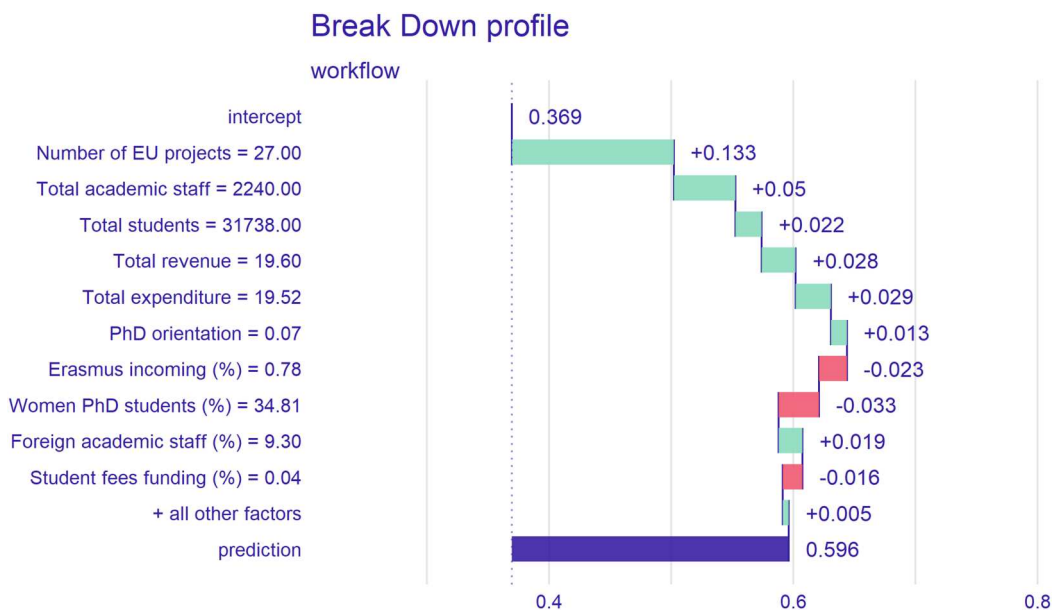
Break Down profile
workflow

intercept	0.369
Number of EU projects = 27.00	+0.133
Total academic staff = 2240.00	+0.05
Total students = 31738.00	+0.022
Total revenue = 19.60	+0.028
Total expenditure = 19.52	+0.029
PhD orientation = 0.07	+0.013
Erasmus incoming (%) = 0.78	-0.023
Women PhD students (%) = 34.81	-0.033
Foreign academic staff (%) = 9.30	+0.019
Student fees funding (%) = 0.04	-0.016
+ all other factors	+0.005
prediction	0.596

Figure 5. Breakdown plot for prediction of English-medium instruction.

of institutions which do not practice EMI: specialised, small, nationally oriented, less research inten-sive, relatively lacking in resources, and undergraduate oriented. In these institutions, national accreditation frameworks, local labour market needs, and the centrality of the national language likely limit both the incentives and the capacity to adopt English-medium instruction. Such insti-tutions likely prioritise domestic missions over international visibility and recruitment, making EMI a weaker strategic fit.

Break Down profile
workflow

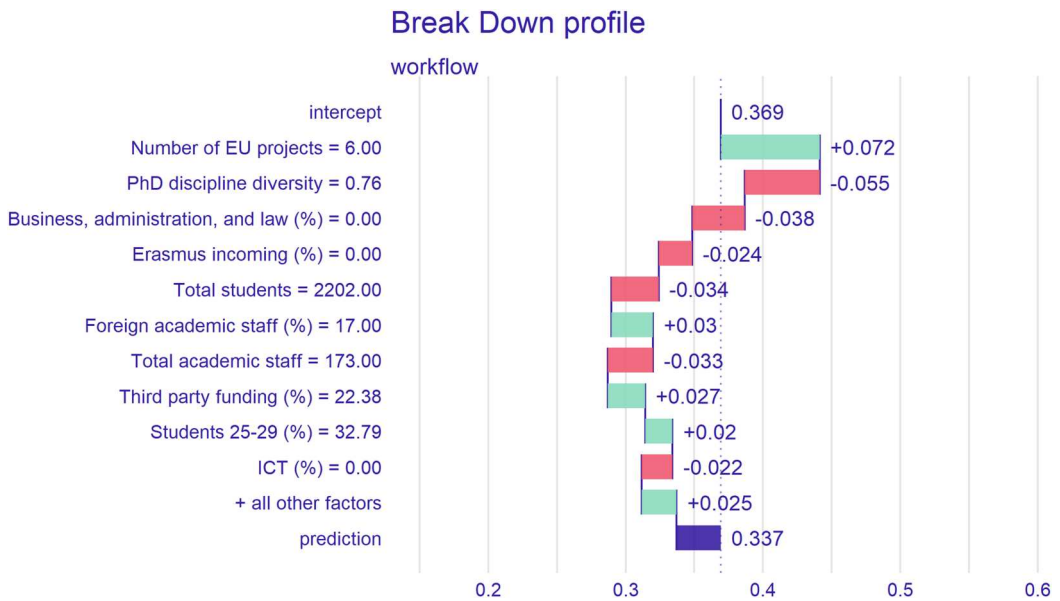|	
intercept	0.369
Number of EU projects = 6.00	+0.072
PhD discipline diversity = 0.76	-0.055
Business, administration, and law (%) = 0.00	-0.038
Erasmus incoming (%) = 0.00	-0.024
Total students = 2202.00	-0.034
Foreign academic staff (%) = 17.00	+0.03
Total academic staff = 173.00	-0.033
Third party funding (%) = 22.38	+0.027
Students 25-29 (%) = 32.79	+0.02
ICT (%) = 0.00	-0.022
+ all other factors	+0.025
prediction	0.337

Figure 6. Breakdown plot for prediction of non-English-medium instruction.

There are several novel insights from these findings. Whilst internationalisation features are strongly suggested by prior theory (e.g. Wilkinson and Gabriëls 2021), the importance of discipline diversity, research, and finance-related features is notable. These are linked to institutional missions, which take an inclusive 'big tent' approach to appeal to a wide range of students; to international cooperation and competition in research-intensive institutions, which favours a common language; and to EMI as a market-responsive product. Perhaps the most interesting finding is also one of the 'least important' (in terms of Gini score), which is that education intensity was largely irrelevant. As EMI is the pedagogic expression of Englishisation (which could be contrasted with academic publishing in English), one might expect education intensity to be important. But it is not. This links back to our opening idea: EMI is not a pedagogic model, but rather an instructional setting; and it is not adopted for pedagogic reasons. Instead, factors related to internationalisation, discipline diversity, research, and finances dominate.

Discussion and implications: isomorphic divergence and 'elitism' versus 'openness'

Our key finding is that EMI is associated with an institutional archetype in Europe: disciplinarily diverse, large, internationally oriented, research intensive, resource rich, and postgraduate focused. The current findings are situated within research on EMI which aims to define, map, and explain this global field. Underlying these discussions are theories on how ideas and practices spread in higher education and ethical concerns on elitism and access.

Prior studies have aimed to map the spread of English-medium instruction worldwide, although with a focus on macro-level trends. An international British Council survey of 55 outer- and expanding-circle countries highlighted the rapid expansion of EMI and observed that it is often promoted top-down by governments (Dearden 2014). This report, and other empirical studies have identified a link between EMI and private education, implying that EMI might act as a mechanism for social and educational stratification (Dearden 2014; Hamid, Jahan, and Islam 2013; Macaro and Akincioglu 2018). A subsequent British Council survey confirmed that EMI was continuing to expand in outer- and expanding-circle contexts, particularly at the undergraduate level, driven by pressures of employability and international competitiveness, and that growth was occurring more rapidly in the private sector (Sahan et al. 2021).

The European context is a part of this global wave, also seeing the rapid expansion of EMI. Between 2013 and 2023, EMI spread out from central and northern Europe to become a pan-European practice, rising from 8,000 ETPs to over 24,000, and the share of institutions offering EMI grew from a quarter to a half (Wingrove et al. 2025). However, Europe differs from global trends with a focus on master's level rather than bachelor's level EMI and, according to our current findings, EMI is equally associated with public and private education. This reflects recent statistical work, which found that being a public or private university did not predict whether an institution offered EMI, except in cases where institutional autonomy was increased, in which case the probability of offering EMI also increased (Wingrove et al. 2024), linking EMI to neoliberal principles of governance.

Underlying these studies are issues related to EMI definitions (Dafouz and Smit 2020; Macaro 2022), typology (Richards and Pun 2023), and operationalisation (Wingrove et al. 2025). Research on EMI typology has identified a range of iterations of EMI, which differ based on purpose, assessment, curriculum model, level, access, and so on (Richards and Pun 2023). The often cited definition of EMI – the use of English to teach academic subjects other than English in countries without English as an L1 (Macaro 2018) – does not reveal the degree of variation we see in EMI worldwide. Our findings have identified that the European variety of tertiary-level EMI is associated with institutions that are disciplinarily diverse, large, internationally oriented, research intensive, resource rich, and postgraduate focused, an iteration that, while reflecting the global trend of EMI expansion, distinguishes the European archetype through the institutional profile in which such growth is

concentrated. These findings echo previous research in Europe. For instance, descriptive statistics finding that EMI is associated with 'larger, PhD-awarding universities with a wide subject range' (Wächter and Maiworm 2014, 51) and with higher ranked universities (Hultgren 2014).

Further studies have found other degrees of variation in EMI as a research field. For instance, finding that EMI is primarily researched by those with an applied linguistics background, rather than non-linguistics who teach the majority of programmes (Macaro and Aizawa 2024) and that EMI journal articles predominantly come from wealthy nations and former colonial powers in the Global North (Hampson, McKinley, and Cargos forthcoming). Taken together: EMI is not a monolith, but rather a differentiated field that generates divisions between participants and institutions within Europe; and positions Europe apart from other regions.

One factor that should be acknowledged is the role of English in European countries compared to countries worldwide. English in the Global South is often a (post-)colonial language, which has led to calls to deploy a decolonial lens in EMI research and practices (Sah and Fang 2025). In Europe the position of English is markedly different; it functions largely as a *lingua franca* and its political dimensions are shaped by debates over integration and national identity rather than a legacy of colonial imposition. Consequently, the propensity for adopting EMI is likely affected by the extent to which English is perceived as an opportunity or as a threat to local linguistic and cultural identity, which differ due to colonial and regional histories.

Although the historical underpinnings are fundamentally different, the global expansion of EMI is an undeniable trend. Explanations for this growth have pointed to (post-)colonialism and globalisation in the global context (McKinley and Galloway 2022) and neoliberal higher education reforms, marketisation, and internationalisation in Europe (Hultgren et al. 2023; Macaro 2018; Wilkinson and Gabriëls 2021). However, macro-level theories of convergence often overlook the institutional response at the meso-level. This study addresses the meso-level directly, arguing that these global trends are built upon a foundation of divergent institutional choices within national contexts.

Succinctly, while global trends point towards inter-country convergence, our findings reveal intra-country divergence. This divergence is visualised in Figure 7, which plots EMI-offering (blue) and non-EMI-offering (grey) institutions in this sample. The institutions that offer EMI are concentrated most in central Europe, which is partially due to the density of institutions in this region. We can also see widespread adoption in Scandinavia, and generally lower adoption South-East Europe. These patterns broadly reflect the latest figures on EMI, which sees EMI practiced broadly in Europe, concentrated in central Europe and Scandinavia, and with South-East Europe trailing (Wingrove et al. 2025). Arguably, one gap in these data is the Balkans, where there are likely region-level factors affecting uptake. Moreover, we should keep in mind that the patterns of adoption are dynamic, not static: as EMI uptake continues to expand, the archetype of an EMI institution is likely to broaden, encompassing a wider range of institutional types.

Crucially, the current findings show that this within-country divergence is not random, but rather associated with certain factors, which can be further understood through the lens of institutional isomorphism. The mechanisms of institutional isomorphism (DiMaggio and Powell 1983) provide a framework for explaining the external pressures that contribute to these patterns. This takes three forms: coercive isomorphism, which results from pressures exerted by other organisations or societal expectations; mimetic isomorphism, which occurs when institutions imitate others perceived as more successful; and normative isomorphism, which stems from professionalisation, including the influence of shared educational backgrounds and networks among professionals. Whilst all these pressures are likely in play in the case of European EMI, we can theorise how they link to our findings.

Mimetic, coercive, and normative isomorphic pressures collectively explain the uptake of EMI, yet also reveal divergent institutional responses based on internal capacity and orientation (Beckert, 2010), suggesting that *selection* mechanisms (i.e. who can respond) are ultimately as consequential as *diffusion* mechanisms (what pressures exist). Concerning mimetic isomorphism, It may be the case that EMI has begun to be associated with more prestigious institutions – larger, well resourced, research intensive, and older – driving others to follow suit. Coercive pressures from national
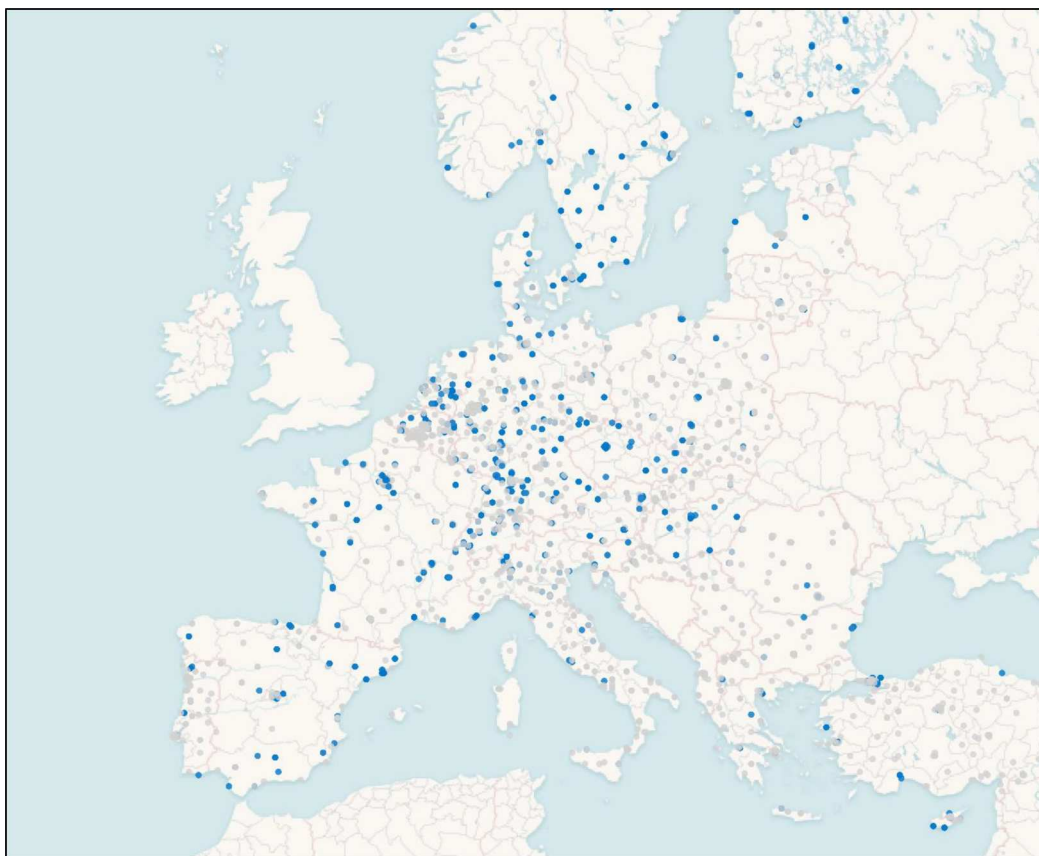
Figure 7. EMI versus non-EMI HEIs in Europe in the academic year 2019/2020.

governments, supranational bodies, and structural influences like the Bologna Process, Erasmus programme, and neoliberal governance principles all incentivise EMI to enhance global competitiveness. However, only certain institutions can effectively respond, like those that are well resourced, international, and already running programmes within disciplines amenable to EMI. Normative isomorphism operates through professional networks and academic staff, who normalise EMI as a legitimate practice; this likely operates through international staff and teaching communities, converging on English in professional practice. Thus, while these external pressures promote convergence, meso-level factors ultimately mediate their impact, leading to within-country divergence.

Consequently, EMI, selectively practiced throughout Europe, risks functioning as a mechanism of elite stratification within academia. Rather than levelling the academic playing field, it may become a marker of prestige and exclusivity, predominantly adopted by large, well-resourced, research-intensive institutions already positioned at the top of national systems. This dynamic exacerbates intra-system inequality, creating a bifurcation between institutions connected to global academic communities, highly-ranked journals, and multinational enterprise through English (Kuteeva and Kaufhold 2024), and those afforded progressively fewer opportunities for such engagement.

If these concerns represent the 'elitism' argument, the opposite side of this debate can be framed as 'openness'. Rather than viewing the English language as erecting barriers to entry, a shared academic language builds bridges. Participation in international research projects and student exchanges – EU funded projects and participation in the Erasmus programme in our data, both associated with EMI – requires participants to converge on a common language.

The function of a *lingua franca* in this context is precisely the opposite of elitism: access is enabled to the widest catchment possible. Moreover, critiques focusing solely on internal academic stratification may overlook an integral function of higher education: the deliberate cultivation of an exclusionary expert class through advanced learning. From an external societal perspective, this makes the university system's role inherently elite; and many students enrol in higher education precisely for this purpose: to gain access to this community and the opportunities it affords. Thus, whilst EMI might reinforce elitism *within* academia it may enable broader access to it from outside.

The central paradox of EMI lies in its capacity to reorient inclusion and exclusion. It does not eliminate barriers but transposes them, shifting the axis of access from the national to the international. Within national systems, our findings confirm that EMI can act as a stratifying force, dividing institutions. Yet, simultaneously, EMI functions as a gateway, offering a wider cohort of international students and scholars access to a global academic community. The critical question, therefore, is not whether EMI is elitist or inclusive (it is both), but how its implementation manages this tension: whether it serves to fortify existing hierarchies or to genuinely democratise participation in global knowledge production. This reframing is essential for formulating thoughtful policy, to which we now turn.

Policy implications

The divergent adoption of EMI and its dual potential for both stratification and openness necessitate nuanced policy responses that move beyond universal mandates. From a national-level policy-maker's standpoint, the heterogenous distribution of EMI across institutions calls for a flexible, differentiated approach rather than one-size-fits-all mandates. For instance, funding formulas based on student graduation rates would favour institutions with stronger capacity to offer EMI, being able to draw on international markets, while placing those that are less able (or willing) to offer such programmes at a disadvantage. Such policies, by implicitly encouraging EMI, may conflict with institutional missions, if they are geared towards specific national industries or professional sectors.

Policymakers should be sensitive to the potential for EMI to result in differential outcomes for students and potentially elite stratification across institutions. To mitigate these risks, policymakers should treat language of instruction as a critical variable in educational planning. Systematic data collection on the differential outcomes of EMI for students, faculty, and institutions is essential to illuminating its role in reinforcing or reducing existing divides.

For institutional leaders, the challenge is to design EMI policies that balance internationalisation goals with the protection of academic freedom, linguistic rights, and equitable access. The case of the Polytechnic University of Milan (Murphy and Zuaro 2021), where a top-down shift to EMI prompted litigation from faculty, serves as a cautionary tale against policies that threaten academic freedom and linguistic rights. A strategic alternative is the development of parallel degree tracks in both English and the local language, ensuring accessibility for domestic students while maintaining international appeal. This approach should be underpinned by robust support systems, ensuring internationalisation complements rather than undermines other institutional missions.

Finally, ethical EMI implementation requires dedicated support structures for all stakeholders. This includes encouraging effective language-subject teacher collaboration (Malmström and Zhou 2025), jointly supporting student language development and academic achievement (Rose et al. 2020), and specific support for foreign marginalised academics (Khuder and Petrić 2023).

Conclusion

This study aimed to identify the meso-level predictors of EMI in European higher education. Applying interpretive machine learning, this study went beyond model performance to theory building, carefully examining the individual predictors. Model performance was assessed using accuracy (84%) and Cohen's Kappa (κ = 6.4), which gave us confidence in the relevance of meso-level features for

predicting EMI. The most important features were: PhD discipline diversity, the number of EU funded projects, the number of academic staff, the share of foreign students, the total number of students, the share of doctoral students, total expenditure, PhD orientation, total revenue, and the share of students in business, administration, and law. Our exploration of all features identified the archetypal EMI-offering institution to be disciplinarily diverse, large, internationally oriented, research intensive, resource rich, and postgraduate focused. Several minor predictors emerged: part-time students, gender balance, and institution age.

Limitations can be considered. Although machine learning uses training and testing sets to avoid overfitting, it should be acknowledged that these data belong to the European context and are not reflective of other EMI contexts worldwide, where we expect substantial differences would be found. This is one of the contributions of the study, that we have identified the archetypical European EMI university, which can be compared to other contexts. Moreover, we have identified some thematic elements which likely transfer, albeit with context-dependent applications, such as the connection between EMI and academic discipline. We can hypothesise that more globalised disciplines are likely to tend towards EMI outside of the European context too, forming part of the international disciplinary community.

A second data limitation is that we see an uneven distribution of feature missingness in our sample (Appendix A), which negatively correlates with the percentage of EMI-offering institutions (-0.55 Pearson's R, suggesting a moderate negative correlation). In other words, participation in data collection for ETER correlates with the offer of EMI, and vice versa. This aligns with our findings, as EMI is associated with international research-intensive institutions, which are more likely to participate in an international research project such as ETER. This exemplifies a methodological issue in large-scale research: the selection bias that arises when a sub-population is inherently less visible and less likely to be captured by standard data collection instruments. Understanding the nature and extent of EMI in these regions is best undertaken by specialised research projects, which seek to understand these hard to reach areas, rather than broad-brush studies such as this one.

In the discussion, we examined how the mechanisms of isomorphic change contribute to divergence in EMI and intersect with issues of elitism. This study underscores the need for policymakers to adopt flexible, context-sensitive approaches to EMI, avoiding policies that inadvertently disadvantage certain institutions. It also calls for closer attention to how language of instruction shapes equity and opportunity within higher education. For institutional leaders, it highlights the importance of strategic planning that aligns EMI initiatives with local contexts, while remaining sensitive to ethical considerations, linguistic diversity, and academic freedom.

Future research directions can be considered. The present study is predictive rather than causal and focuses on institution type, theories of isomorphism, and the implications for inclusivity. However, an intriguing area for future inquiry is the interplay between macro- and meso-level drivers. As an example, prior research identified an interaction between macro-level autonomy and meso-level ownership structure in predicting EMI (Wingrove et al. 2024), suggesting that autonomy acts as a *facilitator* for EMI in response to globalisation pressures. Other relationships hinted at by the current research also merit further investigation, such as links between funding and EMI adoption. In other words: follow the money.

Acknowledgments

Disclosure statement

Funding

ORCID

Peter Wingrove http://orcid.org/0000-0003-2733-0213

References

Allison, P. D. 2001. *Missing Data*. Vol. 136 (Quantitative Applications in the Social Sciences). CA: Sage Publications.

Beckert, J. 2010. "Institutional Isomorphism Revisited: Convergence and Divergence in Institutional Change." *Sociological Theory* 28 (2): 150–166. https://doi.org/10.1111/j.1467-9558.2010.01369.x

Block, D. 2022. "The Dark Side of EMI?: A Telling Case for Questioning Assumptions about EMI in HE." *Educational Linguistics* 1 (1): 82–107. https://doi.org/10.1515/eduling-2021-0007.

Breiman, L. 2001a. "Random Forests." *Machine Learning* 45 (1): 5–32. https://doi.org/10.1023/A:1010933404324.

Breiman, L. 2001b. "Statistical Modeling: The Two Cultures (with Comments and a Rejoinder by the Author)." *Statistical Science* 16 (3): 199–231. https://doi.org/10.1214/ss/1009213726.

Bruni, R., C. Daraio, and D. Aureli. 2021. "Information Reconstruction in Educational Institutions Data from the European Tertiary Education Registry." *Data in Brief* 34:106611. https://doi.org/10.1016/j.dib.2020.106611.

Dafouz, E., and U. Smit. 2020. *ROAD-MAPPING English Medium Education in the Internationalised University*. London: Palgrave Macmillan.

Dearden, J. 2014. *English as a Medium of Instruction-a Growing Global Phenomenon*. London: British Council.

DiMaggio, P. J., and W. W. Powell. 1983. "The Iron Cage Revisited: Institutional Isomorphism and Collective Rationality in Organizational Fields." *American Sociological Review* 48 (2): 147–160. https://doi.org/10.2307/2095101

Grimmer, J., M. E. Roberts, and B. M. Stewart. 2021. "Machine Learning for Social Science: An Agnostic Approach." *Annual Review of Political Science* 24 (1): 395–419. https://doi.org/10.1146/annurev-polisci-053119-015921

Hamid, M. O., I. Jahan, and M. M. Islam. 2013. "Medium of Instruction Policies and Language Practices, Ideologies and Institutional Divides: Voices of Teachers and Students in a Private University in Bangladesh." *Current Issues in Language Planning* 14 (1): 144–163. https://doi.org/10.1080/14664208.2013.771417

Hampson, T, J. McKinley, and K. Cargos. Forthcoming. "MI: Who's Studying Whom? A Critical Bibliometric Analysis of Authorship and Citation Patterns in English Medium Instruction Research." *Higher Education*.

Hultgren, A. K. 2014. "English Language use at the Internationalised Universities of Northern Europe: Is There a Correlation between Englishisation and World Rank?" *Multilingua-Journal of Cross-Cultural and Interlanguage Communication* 33 (3-4): 389–411. https://doi.org/10.1515/multi-2014-0018

Hultgren, A. K. 2024. "English-medium Instruction in Higher Education throughout Europe." In *The Routledge Handbook of English-Medium Instruction in Higher Education*, edited by K. Bolton, W. Botha, and B. Lin, 81–95. London: Routledge.

Hultgren, A. K., M. Nao, P. Wingrove, D. Yuksel, and B. Zuaro. 2023. "New Insights into the Trend towards English as a Medium of Instruction in European Higher Education through Transdisciplinary Participation: Poikkitieteellinen Yhteistyö ja sen Tuomat Uudet Näkökohdat Korkeakouluopetuksen Englanninkielistymiseen Euroopassa." *AFinLAn Vuosikirja* 25:318–331.

Hultgren, A. K., and R. Wilkinson. 2022. "New Understandings of the Rise of English as a Medium of Instruction in Higher Education: The Role of key Performance Indicators and Institutional Profiling." *International Journal of the Sociology of Language* 2022 (277): 47–59. https://doi.org/10.1515/ijsl-2021-0082

Khuder, B., and B. Petrić. 2023. "Intersectionality of Marginalisation: Eal academics in Exile Writing for International Publication." *Journal of Multilingual and Multicultural Development* : 1–14. https://doi.org/10.1080/01434632.2023.2170387.

Kuhn, M., and J. Silge. 2022. *Tidy Modeling with R*. Sebastopol, CA: "O'Reilly Media, Inc.".

Kuteeva, M., and K. Kaufhold. 2024. "An 'E'for 'Elite'in EMI? Global, Local and Elite Dimensions in the Promotion of English-Medium University Programmes." *Journal of Multilingual and Multicultural Development* : 1–17. https://doi.org/10.1080/01434632.2024.2393707

Landis, J. R., and G. G. Koch. 1977. "An Application of Hierarchical Kappa-Type Statistics in the Assessment of Majority Agreement among Multiple Observers." *Biometrics* 33 (2): 363–374. https://doi.org/10.2307/2529786.

Macaro, E. 2018. *English Medium Instruction*. Oxford: Oxford University Press.

Macaro, E. 2022. "English Medium Instruction: What Do We Know So far and What Do We Still Need to Find out?" *Language Teaching* 55 (4): 533–546. https://doi.org/10.1017/s0261444822000052.

Macaro, E., and I. Aizawa. 2024. "Who Owns English Medium Instruction?" *Journal of Multilingual and Multicultural Development* 45 (10): 4037–4050. https://doi.org/10.1080/01434632.2022.2136187

Macaro, E., and M. Akincioglu. 2018. "Turkish University Students' Perceptions about English Medium Instruction: Exploring Year Group, Gender and University Type as Variables." *Journal of Multilingual and Multicultural Development* 39 (3): 256–270. https://doi.org/10.1080/01434632.2017.1367398

Macaro, E., S. Curle, J. Pun, J. An, and J. Dearden. 2018. "A Systematic Review of English Medium Instruction in Higher Education." *Language Teaching* 51 (1): 36–76. https://doi.org/10.1017/S0261444817000350

Madley-Dowd, P., R. Hughes, K. Tilling, and J. Heron. 2019. "The Proportion of Missing Data Should Not Be Used to Guide Decisions on Multiple Imputation." *Journal of Clinical Epidemiology* 110:63–73. https://doi.org/10.1016/j.jclinepi.2019.02.016

Malmström, H., and S. Zhou. 2025. "Language – Subject Teacher Collaboration in English-Medium Higher Education: Current Practices and Future Possibilities." *RELC Journal*. https://doi.org/10.1177/00336882241313234.

McKinley, J., and N. Galloway. 2022. *English-medium Instruction Practices in Higher Education: International Perspectives*. London: Bloomsbury Publishing.

Molnar, C. 2022. *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable*. Munich: Leanpub. https://christophm.github.io/interpretable-ml-book/.

Murphy, A., and B. Zuaro. 2021. "Internationalization vs Englishization in Italian Higher Education: Reframing the Issue." In *Englishization of Higher Education in Europe*, edited by R. Wilkinson, and R. Gabriels, 163–188. Amsterdam: Amsterdam University. https://doi.org/10.1515/9789048553914-010.

Painsi, P. W., K. H. Leitner, and R. Rybnicek. 2024. "The Impact of Institutional Autonomy on Higher Education Institutions in Europe." *Science and Public Policy* 52 (2): 236–253.

Pecorari, D., and H. Malmström. 2025. "English Medium Instruction and English for Specific Purposes." In *The Handbook of English for Specific Purposes*, edited by S. Starfield and C. Hafner, 167–184. Hoboken, NJ: Wiley. https://doi.org/10.1002/9781119985068.ch9.

Richards, J. C., and J. Pun. 2023. "A Typology of English-Medium Instruction." *Relc Journal* 54 (1): 216–240. https://doi.org/10.1177/0033688220968584

Rose, H., S. Curle, I. Aizawa, and G. Thompson. 2020. "What Drives Success in English Medium Taught Courses? The Interplay between Language Proficiency, Academic Skills, and Motivation." *Studies in Higher Education* 45 (11): 2149–2161. https://doi.org/10.1080/03075079.2019.1590690

Sah, P. K., and F. Fang. 2025. "Decolonizing English-Medium Instruction in the Global South." *Tesol Quarterly* 59 (1): 565–579. https://doi.org/10.1002/tesq.3307

Sahan, K., A. Mikolajewska, H. Rose, E. Macaro, M. Searle, I. Aizawa, S. Zhou, and A. Veitch. 2021. *Global mapping of English as a medium of instruction in higher education: 2020 and beyond*. London: British Council.

Stigmar, M. 2016. "Peer-to-peer Teaching in Higher Education: A Critical Literature Review." *Mentoring & Tutoring: Partnership in Learning* 24 (2): 124–136. https://doi.org/10.1080/13611267.2016.1178963.

Wächter, B., and F. Maiworm. 2014. *English-taught Programmes in European Higher Education: The State of Play in 2014*. Bonn: Lemmens Medien GmbH.

Wagner-Schuster, D., B. Lepori, A. Lambrechts, and G. Zahradnik. 2023. ETER – Full dataset (csv) [Data set]. Zenodo. https://doi.org/10.5281/zenodo.8074203.

Wilkinson, R. 2013. "English-medium Instruction at a Dutch University: Challenges and Pitfalls." *English-medium Instruction at Universities: Global Challenges* 324 (10.21832): 9781847698162-005.

Wilkinson, R., and R. Gabriëls, eds. 2021. *The Englishization of Higher Education in Europe*. Amsterdam: University Press. https://doi.org/10.5117/9789463727358.

Wingrove, P., B. Zuaro, M. Nao, D. Yuksel, L. Littvay, and A. K. Hultgren. 2024. "University Autonomy Is a Predictor of English Medium Instruction in European Higher Education." *Higher Education* 90:487–520. https://doi.org/10.1007/s10734-024-01333-8.

Wingrove, P., B. Zuaro, D. Yuksel, M. Nao, and A. K. Hultgren. 2025. "English-Medium Instruction in European Higher Education: Measurement Validity and the State of Play in 2023/2024." *Applied Linguistics*. https://doi.org/10.1093/applin/amaf020.

Wright, M. N., and A. Ziegler. 2017. "ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R." *Journal of Statistical Software* 77 (1): 1–17. https://doi.org/10.18637/jss.v077.i01

Yuksel, D., M. Nao, P. Wingrove, B. Zuaro, and A. K. Hultgren. 2024. "Macro-level Language Policy and Planning to Promote and Maintain English-Taught Programmes in Turkish Higher Education: A Process Tracing Perspective." *Current Issues in Language Planning* : 1–23. https://doi.org/10.1080/14664208.2024.2400836