1. **Abstract (what we want to do)**

In this study we ~~can~~ observe data related to students of 15 years of age coming from European countries ~~among which there’s~~ with a significant number of immigrant students. How do immigrant students face up when compared to native students? Which are the main differences in their scholastic, familiar, psychological characteristic? Which features are most important when it comes to their scholastic success? How can we take example from the best European countries with respect to integration to help these students?

High School years are a fundamental stage of development for everyone, and it is in the interest of every country to ensure that all students have access to the best possible education and are well integrated into the social system.

We analyzed data related to students of 15 years of age coming from European countries which have a significant number of immigrant students. We observe that immigration is a huge factor limiting ~~many~~ foreign students from achieving the same ~~scholastic~~ results of their native peers. This trend is common ~~to all~~ to most countries across Europe, ~~but some of them show significantly larger gaps with respect to the others~~ but some worse than others.

In this study we aimed at investigating quantitatively the reasons behind these differences.

1. **Introduction/objectives (how we want to do it)**

Our dataset is based on some questionaries’ answers of students, their parents, and their school’s staff, regarding both their scholastic knowledge and social and psychological conditions.

After exploring the dataset and analyzing it through MANOVA tests and Clustering methods, we focused on students’ scores in Math and Reading. Indeed, native and immigrant students show consistent differences in term of their results.

Then through linear models, we identified the main covariates responsible of the students scores~~, such as his socioeconomic state, the amount of time he dedicates to study, his class size and some others, besides his hard work and study.~~ ~~These results suggested us some ways to help the lagging students, who most of the times were the immigrant ones, understanding the points that could be improved.~~ to understand where to act to help struggling students, often immigrant, in the most effective way.

Finally, through Linear Mixed Models we clustered schools based on their contribution to the scores and using Multinomial Logistic Regression we identified the key characteristic which allows the “better” schools to help their students.

1. **Materials & Methods (how we did it)**

* Data selection

Starting from the data of Pisa’ OECD program of 2018, which contains student and school answers to standardized questionnaires, we restricted our analysis to 10 selected countries which had a sufficient sample of immigrant students: Austria, Belgium, Denmark, Germany, Great Britain, Italy, Luxemburg, Spain, Sweden, Switzerland.

Chart, sunburst chart

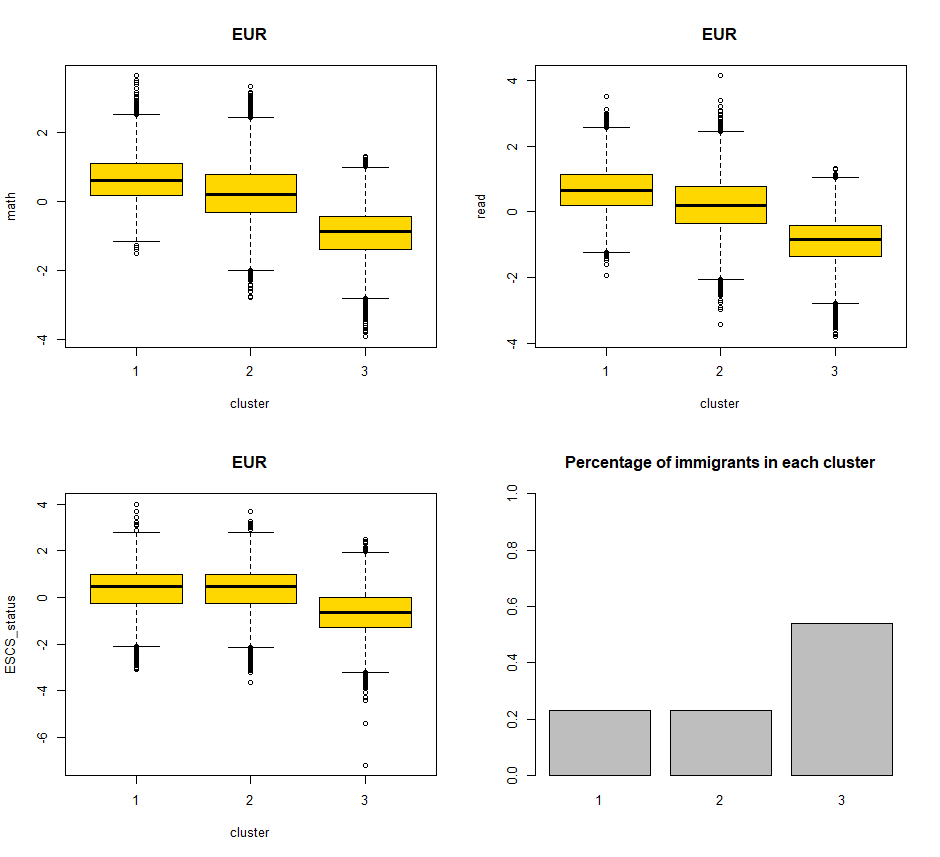
Description automatically generated

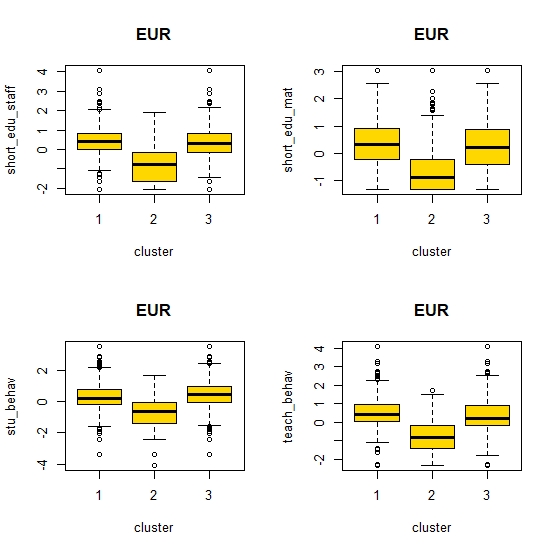
~~Since the number of covariates available was more than a thousand (one for each questionnaire answer), we selected the most relevant ones related to our focus on immigration.~~ Some questionnaire answers where already grouped in aggregated features, ~~of the questions answered by the students were already aggregated by Pisa’s group and resulted to be the most complete.~~

The features we selected range over many fields: ESCS status (index of economic, social, and cultural status), Immigration status, teacher support, sense of belonging, class size, at home weekly learning time, etc.

* Clustering

~~We first qualitatively visualised our dataset trough clustering.~~ Investigating the dataset through k-means ~~we came up with three balanced clusters which are not perfectly divided but seem~~ we identified 3 clusters of students.





The cluster with the worst math and reading scores contains 54% of the total number of immigrants, while the best contains only 23% of them.

Moreover, the clusters show that most immigrant students attend poorer schools with a shortage of material and staff, and a school climate that hinders learning.

* MANOVA

~~We then used MANOVA to get statistical evidence for our hypothesis of there being differences between native and immigrant students in math and reading scores~~.

To confirm the qualitative results observed through clustering we performed MANOVA tests on the main features of interest.

For all countries (except Great Britain) we found statistical evidence at level 95% that immigrant students have lower mean scores in both Math and Reading.

By ranking the differences in scores, we see that Great Britain is first while Denmark is last.

Similar differences are found also for other features such as ESCS status, sense of belonging in school, grade repetition percentages.

On the other hand, weekly times spent studying are consistently higher for immigrant students compared to native, which excludes lack of commitment as the motivation behind lower scores.

* Linear models

~~After assessing the presence of differences in scores between immigrants and natives we tried to understand which features are most important for a student’s scholastic success, to find the optimal way for schools to help lagging students and to have a better integration system.~~

~~By focusing~~ to find the features mainly responsible of scholastic success we focused on three datasets: our complete dataset, Great Britain as an example of a country with good integration, and Denmark as one with more integration issues.

For immigrant students some of the main negative regressors are: not speaking the country’s language at home, having parents who are not highly educated and being bullied at school.

Moreover, schools with a shortage of educational material and staff influences their students’ achievements, especially if they are immigrant since they might need additional attention from teachers.

From the separate analysis of Great Britain, we noticed that ‘immigration’ and most of its interaction effects were not significant anymore, as expected. The key factors are schools’ resources and ESCS status.

On the contrary, in Denmark being an immigrant has a strong negative effect, worsened by the interaction with other covariates such as the student-teacher ratio in schools and the emotional status of the students.

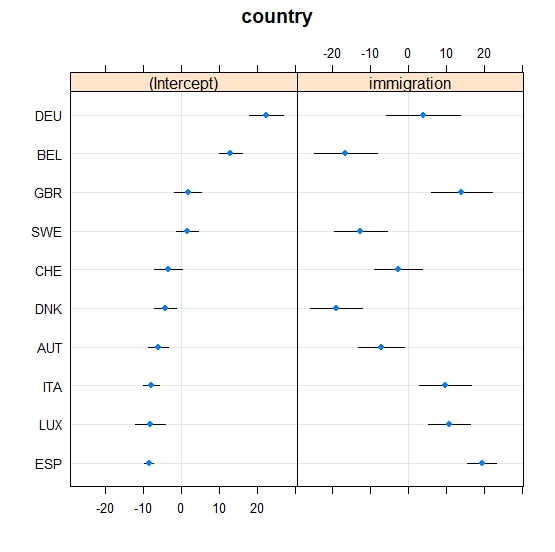
* LMM

Since our data offers some grouping possibilities (by country and by schools) we extended our linear models into linear mixed models aggregating the observations firstly by countries, adding a random slope associated with the variable immigration.

Secondly, we grouped the datasets of Great Britain and Denmark by schools, to observe what do schools in these countries do differently to help their students through Multinomial Logistic Regression in the next section.

The random effect given by the country grouping is significant and adding the immigration random slope increases its relevance. The Percentage of Variance explained by the Random Effect (PVRE) goes from 2.3% to 4.4% for math and from 1.5% to 4.3% for reading.

Chart

Description automatically generated

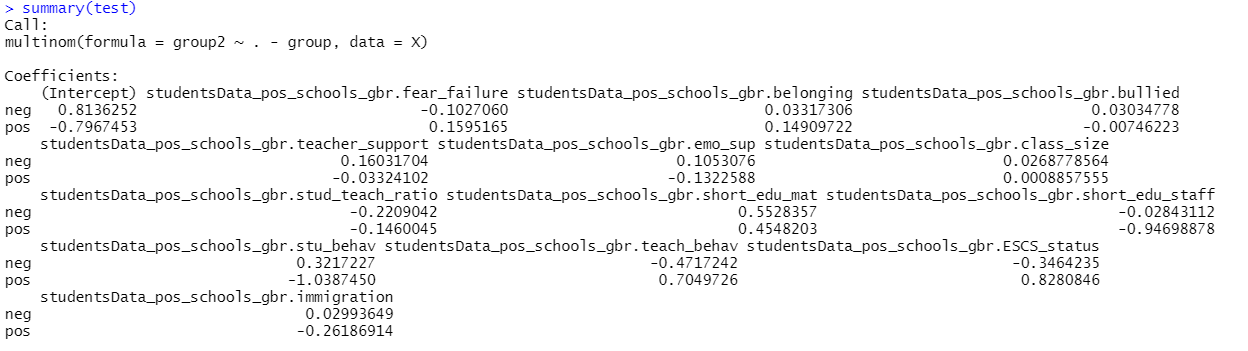
By plotting the random coefficients of the two models, it’s clear how being an immigrant living in Great Britain has a positive effect on the school performances while living in Denmark has a negative one with respect to the mean.

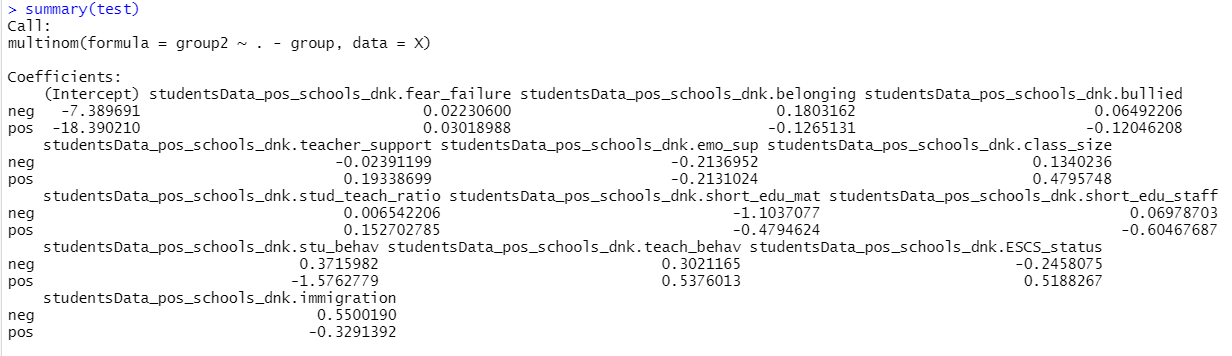
* Multinomial logistic regression models

We implemented two mixed models for the schools of GBR and DNK (see previous section), then for each we divided the schools in three clusters: those with a positive intercept, those with no significant effect, and those with negative intercept.

We then identified the different characteristics of the “better” and “worse” schools through Multinomial Logistic Regression.

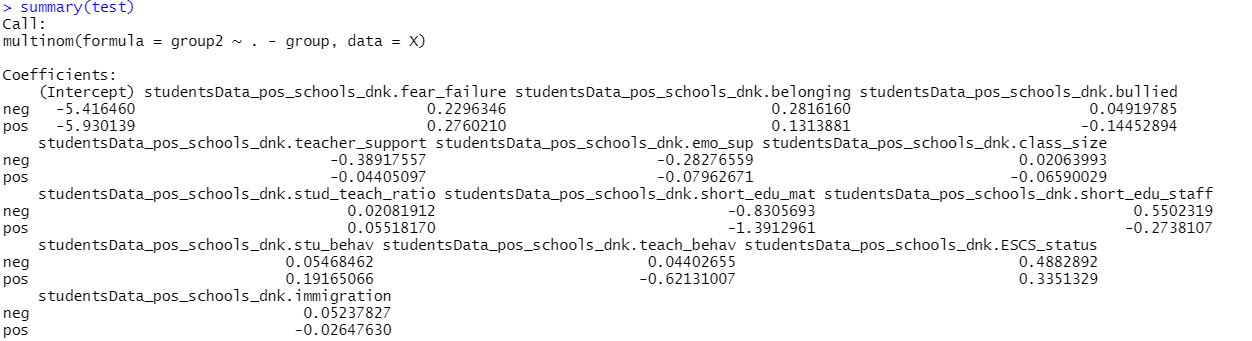
Indeed, the output of these models assign at each covariate a coefficient representing their contribution in increasing the probability of being assigned to a better or worse school.





Text

Description automatically generated



We notice that in Great Britain the better and worse schools have similar characteristics, except for higher ESCS status of the students and larger educational staff, which might give more attention to students’ individual issues.

This is also true in Denmark, but in the worse schools there is also a relevant worsening in the school climate and, as expected, being an immigrant student increments the probabilities of being assigned to the worse schools. Indeed in Denmark the worse schools have a higher percentage of immigrant students.

1. Results ( what were the results)
2. Conclusions (What do we learn from the results)