

---

# Minorities' results on the Romanian Baccalaureate

---

**Orsolya-Renáta Szőcs**  
Matriculation nr. 6158974

**Alpár Cseke**  
Matriculation nr. 6159070

{orsolya.szocs, alpar.cseke}@student.uni-tuebingen.de  
<https://github.com/alparius/bac-stat>

## Abstract

The topic of the Romanian Baccalaureate (secondary school graduation exam) instigates several assumptions regarding the difference between the results of the country's ethnolinguistic groups and that of the main population. After having scraped the last few years' results, our goal is to discuss and visualise different aspects of the gathered data to find out whether the common beliefs have a basis, and to get a better understanding of the potential underlying issues.

## 1 Introduction

### 1.1 The Romanian Baccalaureate

The Romanian Baccalaureate [1] is the exam taken by students in Romania when they graduate high school. For the most part, the baccalaureate consists of three written exams. Each written exam is graded on a scale of 1 to 10 with 10 being the best possible grade. For students to pass the exam as a whole they need to obtain a minimum of 5 on each written module and a final average of at least 6.

Romanian language and literature is one of the three written modules. This exam has to be taken by every exam participant. Moreover, each student must take a mandatory and an elective exam based on their class's study profile. In addition to these three exams, many students who belong to an ethnic minority group are also studying their native language and literature, thus must take a fourth exam in this subject.

### 1.2 Minorities of Romania

Romania is home to several ethnolinguistic groups who can study their native language and literature in school. The most prevalent minority group in Romania are the Hungarians with 6.3% of the population, according to the country's last census in 2011 [2]. Other linguistic minorities include the Ukrainian (0.24%), German (0.13%) and Turkish (0.13%) groups.

### 1.3 Motivation

There are several assumptions that people in Romania hold about the results of the minority groups on the baccalaureate. Some of these suppositions are the following:

- The pupils belonging to minority groups receive worse final grades and in particular, worse grades in Romanian language and literature. The general argument for the first part of this presumption is that since the minorities need to take an additional exam, they have less time to prepare. On the other hand, people claim that as Romanian is not the native language of the minority groups, they likely have lower grades in this subject.

- Better minority students are less affected by their lack of Romanian linguistic knowledge on the Romanian literature exam.
- Students who are part of a minority have a worse chance of successfully increasing their Romanian literature grade through the grade appeal process as they get penalised for their grammatical errors that are specific to non-native Romanian speakers.

It is worth bearing in mind that these suppositions merely rely on anecdotes and personal observations. Our goal is to challenge these assumptions by utilising the available data.

## 2 Methodology

Our language of choice for manipulating the data was *Python*, together with its staple data stack (i.e. *pandas*, *numpy* and *matplotlib*). The experiments themselves were carried out in *Jupyter Notebook* to facilitate a step-by-step workflow and make results reproducible.

### 2.1 Data Acquisition

The underlying dataset was acquired from the official site of the examination [3]. Only the last three years (2019, 2020, 2021) were available. Earlier results were removed from the site since those were not anonymised.

The data was scraped with the *Selenium* library [4], which utilises a headless browser to render the target webpages using a stripped-down version of the browser’s engine. Static scraping methods did not prove to be successful due to the *JavaScript* code present in several of the table’s cells.

The scraping itself turned out to be times of magnitudes slower than initially expected, therefore the script was extended for parallel execution: to save data to files in smaller batches and to always check for already collected chunks by other processes.

Furthermore, it must be noted that the scraping process abode by the webpage’s *robots.txt* ruleset: the resource was not disallowed and we respected the 10 second request frequency (although the long running time was also a limitation of our code). The total process involved not more than 40 000 URL requests spread across multiple days.

## 3 Experiments and results

In this section we examine the results of the baccalaureates held in 2019, 2020 and 2021. Since the difficulty level of the individual exam modules varies each year, we believe that through combining the results from the three years we obtain data that is more representative of the general baccalaureate exam.

We denote the group of students who had to take only three written exams as *R*. Those studying one of the minority languages are treated as a single group *M* because they share the common trait of having four exams instead of three. It must be noted that their actual distribution is similar to the one presented in subsection 1.2, with Hungarians accumulating to 5.3% of exam participants and all of the other minorities totalling up to 0.7%.

Table 1: Number of participants at the Romanian Baccalaureate

Student group	Size (2019)	Size (2020)	Size (2021)	Total size
<i>R</i>	121 693	140 315	120 472	382 480
<i>M</i>	6 826	7 351	6 381	21 528

Table 2: Average final grades at the Romanian Baccalaureate

Student group	Average (2019)	Average (2020)	Average (2021)	Total average
<i>R</i>	7.007	6.820	7.030	6.946
<i>M</i>	6.868	6.994	7.013	6.958

Table 1 shows the number of the students not disqualified from any of the exams (i.e. with valid grades). Table 2 displays the average final grades of the two groups in each year. Although we can not tell whether or not the grades of  $M$  would be greater if they only had to take three exams, we observe that the difference between the average final grades of  $R$  and  $M$  is very small. Therefore, based on these three years' data,  $M$  does not seem to be at disadvantage when considering only the final grades.

The distribution of the final grades in Romanian language and literature is shown in Figure 1. A notable aspect of this distribution is that there are relatively few grades between 4 and 5 and many grades between 5 and 5.5. A probable explanation for this is that teachers correcting the exam might be less strict when the student's grade would otherwise fall right below the minimum passing grade.

The plot shows that group  $M$  achieves worse grades than  $R$  on this particular exam. The average grade in this subject for  $R$  is 7.060, while for  $M$  it is 5.744. Thus, on average, group  $R$  achieves a grade greater by 1.316, which equates to a difference of 14.62%.

It is worth comparing  $M$ 's grade in their native language and literature to  $R$ 's Romanian literature grade.  $M$ 's average native language grade is 7.630, which is greater by 0.570 than the average Romanian grade of  $R$  and by 1.886 than the average Romanian grade of  $M$ . Yet, this greater native literature grade does not fully account for how group  $M$  makes up for their lesser Romanian literature grades to have an almost even average final grade.

Figure 2 shows that the rest of the discrepancy between the two groups lies in their specialisation-specific grades. Proportionally to their sizes,  $M$  achieves considerably more grades in the 6 to 9 grade buckets, summing up to a 0.294 higher average grade in the two specialization-specific exams. Specifically,  $R$ 's average is 6.888 as opposed to  $M$ 's 7.182.

This difference is not negligible, since the mandatory and elective exams of the specialisation should have very little to do with the participants' native languages. The phenomenon could have several different explanations: minority schools might better prepare students for their study profile exams, or the minority exam graders could be less strict.

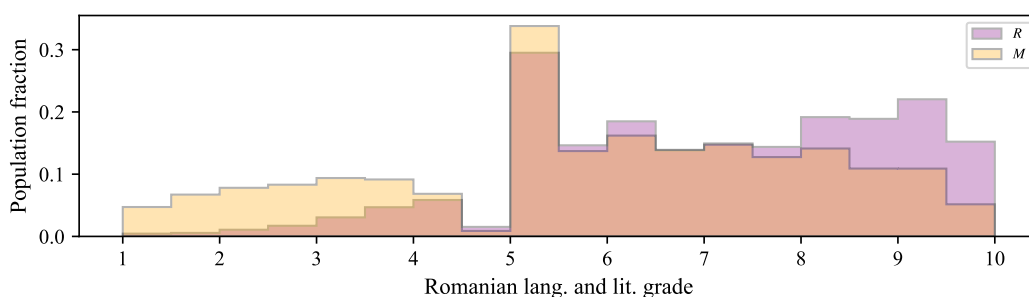


Figure 1: Distribution of the Romanian language and literature grades

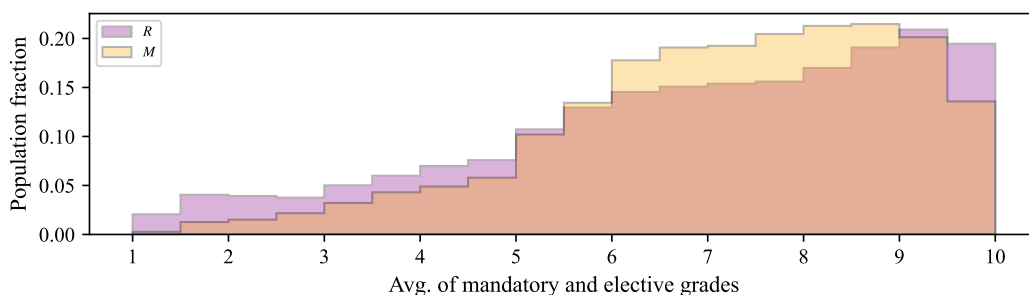


Figure 2: Distribution of the average of the mandatory and elective subject grades

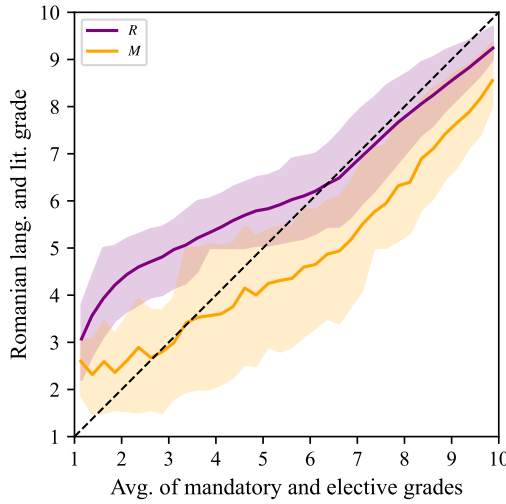


Figure 3: Expected Romanian language and literature grades given the average of the specialisation’s mandatory and elective grades

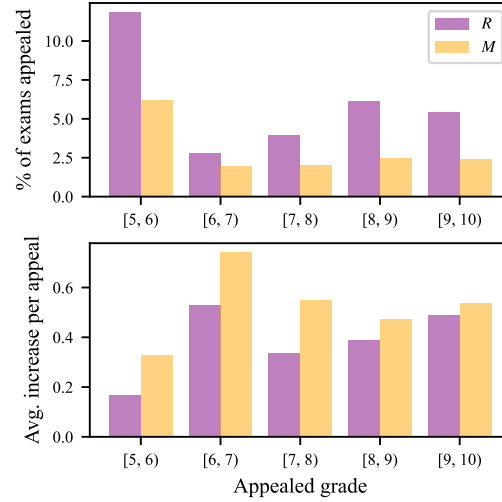


Figure 4: Percentage of students in each grade group submitting a grade appeal for their Romanian lang. and lit. exam (top). Average increase in grade after the appeal (bottom)

Figure 3 is achieved by separating the examinees into 36 groups based on the average of the grades specific to their study profiles and computing the average Romanian language grade for each group. Additionally, the plot also shows the area bounded by the 25<sup>th</sup> and 75<sup>th</sup> percentiles.

We can observe that in the case of both *M* and *R* better specialisation grades generally lead to better grades on the Romanian exam. Supporting the assumption that better *M* students are at a lesser disadvantage at the Romanian literature exam, we can notice that the difference does tighten between the Romanian grades of *M* and *R* when their average specialisation grade is above 8.

Our last assumption underlined in subsection 1.3 is regarding the improper treatment of Romanian language and literature grade appeals submitted by *M* members. What the top part of Figure 4 reveals is that only a considerably smaller percentage of *M* requests a grade appeal in this exam module. Presumably, they require more confidence in a successful re-evaluation. This is a feasible explanation to the bottom plot of the figure, namely that *M* receives on average a higher increase in their final grade from their grade appeals.

## 4 Conclusion

The work on the subject is far from conclusive. Having not done actual hypothesis testing or causal inference, we can not claim any quantifiable results over our initial assumptions. More so, this paper is intended as a discussion starter in the grand topic of examination fairness across the diverse minority groups of today’s countries.

## References

- [1] Ministry of Education. *Metodologia de organizare și desfășurare a examenului de bacalaureat*. (2011). URL: <http://static.bacalaureat.edu.ro/2021/docs/MetodologieAnexa2.pdf> (visited on: 2022. Feb. 6.)
- [2] Romanian National Institute of Statistics. *Population and housing census - 2011*. Volume II: Stable (resident) population - ethnic and confessional structure. (2011). Chapter 10: Population by ethnicity and mother tongue - counties. URL: <https://www.recensamanromania.ro/rpl-2011/rezultate-2011/> (visited on: 2022. Feb. 6.)
- [3] Ministry of Education. *Bacalaureat 2021 - Rezultate*. (2021). URL: [http://static.bacalaureat.edu.ro/2021/rapoarte/rezultate/alfabetic/page\\_1.html](http://static.bacalaureat.edu.ro/2021/rapoarte/rezultate/alfabetic/page_1.html) (visited on: 2022. Feb. 6.)
- [4] Selenium. URL: <https://www.selenium.dev> (visited on: 2022. Feb. 6.)