EXAM RESULTS AT THE END OF HIGH SCHOOL

Insights from microdata on exams in Portugal, 2008-2022

NOVA IMS Information Management School

Joule data journalism courses for higher education

Co-funded by the Erasmus+ Programme of the European Union

Introduction

The remote inspiration for this work was Biology Professor Jorge Salcedo, at Liceu Camões. When students were protesting against the new exam system (Exame Geral de Acesso) e offered another angle, saying that we were in a demanding school and would only benefit from national exams, contrary to students attending less demanding schools.

This was many years ago. In the last decades data became more readily available, with newspapers such as Público and Observador publishing the (in)famous rankings. And Júri Nacional de Exames (National Exam Jury) now publishes one Access database file per year with Information on exams.
https://www.dge.mec.pt/relatoriosestatisticas-0

As my daughter Carolina and my son Ricardo were finishing high school, my interest was revived. Last year I presented an analysis comparing internal school results with exam grades to their former school's "Conselho Geral".

Methodology

Using Python, all files available (2008-2022) were combined in order to facilitate analysis (project available on Github - link in the corner).

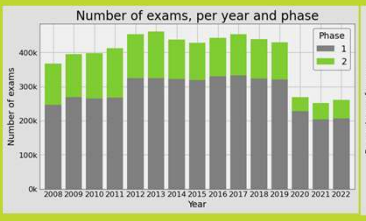
This resulted in a dataset with 5.900.488 exam results that correspond to dozens of disciplines, both Phase 1 and 2 of the exams, from all schools that take these exams. These include the 14 exams taken by Carolina and Ricardo.

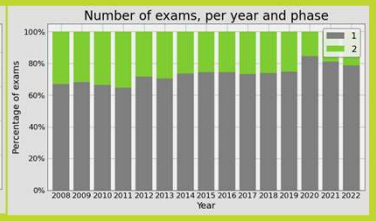
We are using the actual, complete data (census), not samples that might not be representative.

We have Information on the school, discipline, gender and age of the student (but no name or any form of identification), phase, reason why the exam is taken, exam grade, internal grade and final grade. No other sources were used, other than these ENES databases.

Info: Characterizing the exams

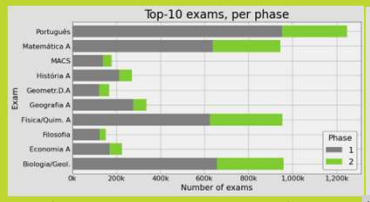
Number of exams

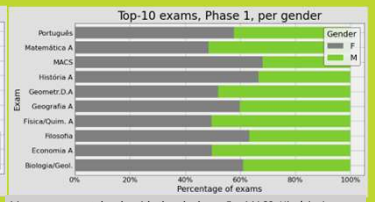




There's an apparent trend of small but consistente growth in the number of exams taken.
Then in 2002, the number of exams fell abruptly, and kept low for the next two years.
In 2020 the number of exams required was reduced and exams to improve grades from the previous year were not allowed.
Both phases have less exams, but for Phase 2 the drop is more pronounced.

Looking at the exams

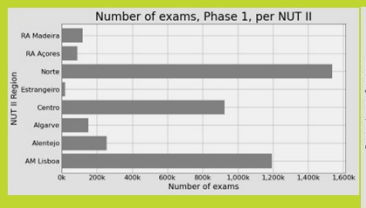


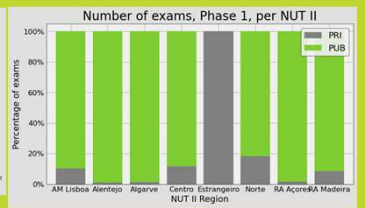


Português is clearly the most common exam in Phase 1, followed by Física e Química A, Biologia e Geologia e Matemática A.
On Fase 2, the four have roughly the same number of exams.

More exams are taken by girls than by boys. For MACS, História A, Filosofia and Biologia/Geologia, more than 60% of the exams are done by girls, and for Português and Geografia A, it's also close to 60%. The number is balanced for Matemática A, Geometria Descritiva A, Física e Química A and Economia A.

Regional distribution

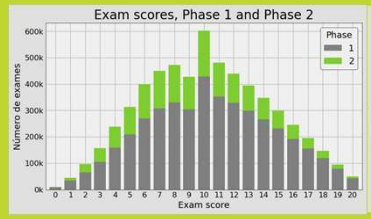


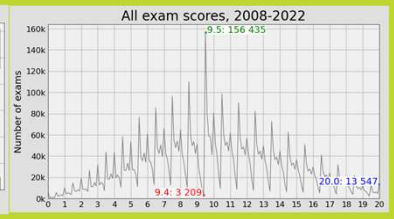


Regions are very diverse in the number of exams taken, and also in the proportion of exams coming from private schools (in some, it's almost zero).

Info: Results

Number of exams

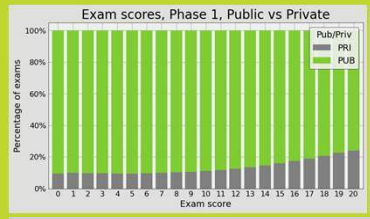


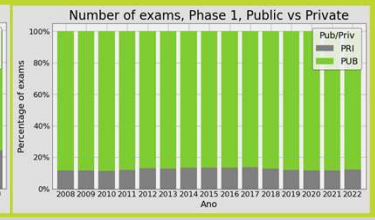


The representation of the results (rounded to integer) resembles a normal distribution, symetric and centered at 10, with extreme values being rare.
Also note how Score=9 is below expected.

Using the actual values, the normal distribution seems mixed with an ECG! Scores with .4 are rounded to .5, which rounds then to the next integer. This difference is more evident and more dramatic at the 9.4 level (fail), while 9.5 means approve.
9.4 is is extraordinarily rare; to its right, 9.5 is the most common grade.

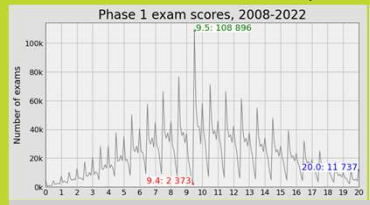
Results, public vs private

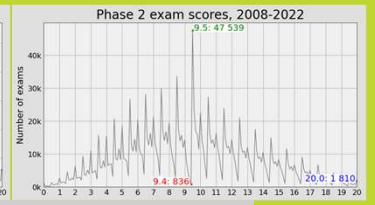




We don't see a large variation over time of the percentage of exams coming from public compared to private schools.
We see that private schools are more represented in the higher grades than they are in the lower grades, and this increases consistently.

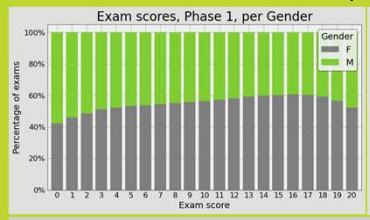
Results, Phase 1 vs Phase 2

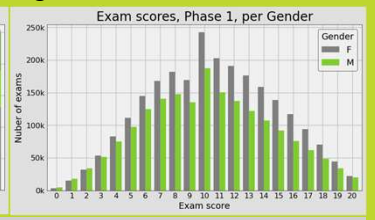




When we break down the distribution of grades by phase, as expected, Fase 2 results lean more to the left.

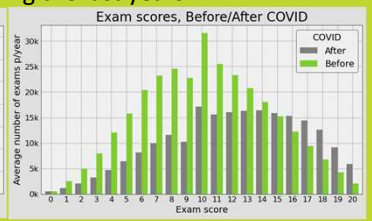
Results, per gender





The distribution by gender shows that boys are less represented in the intermediate grades, and more in both extremes, positive and negative.

Results, comparing the last years



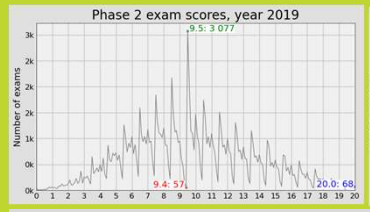
When we compare the last 6 years, it's evident that there's a huge difference between the first 3 and the last 3.
Results from 2020-2022 are no longer centered at 10 and lean significantly to the right.
If we aggregate and average results before and after COVID, it 's evident that higher results became significantly more popular.

The COVID miracle

2019 vs 2020, Phase 1

In 2020, the changes in structure and criteria that were established, attributed to the limitations due to COVID restrictions, had a tremendous impact on the results.
The distribution that used to by symetrical and centered around 9.5 was in 2020 quite assymetrical and leaned towards the higher grades.
In all previous and following years, 9.5 was always the most common result. In Phase 1, 2020, the most common result was 20.0.
In 2019, out of 321.196 phase 1 exams, there were 618 grades of 20.0; in 2020, out of 227.530 exams (-29%), this number jumped to 3.808 (6x).

2019 vs 2020, Phase 2



In Phase 2, 9.5 recovered its position as the most common result.
20.0, no longer the most common, remained a very frequent result even in Phase 2.

The road ahead

Regarding education

- The trend in Education is for exams to account for less, in terms of access to higher education;
- There's now less transparency regarding this information; since 2020 we no longer have information on internal grades in the files;

Further research (WIP)

- Until 2019 we have info on internal grades; we're studying the factors that explain internal grade, using the ENES DB;
- Bringing "outside data" from census, university enrolment, etc, would enable do deepen and extend this analysis;
- Information on internal grades for 2020 and further;
- If we could somehow relate several exams by the same student, also connecting 9th and 12th grade exams, that would open up a new world of possibilities to investigate;

This will take part of the study by Prof. Gil Nata e Tiago Neves, and try to look at it from other angles:
https://www.cnedu.pt/content/iniciativas/seminarios/2015_04_07_CNE_evora.pdf

About this poster

First and foremost, I dedicate this work to **Carolina e Ricardo**, my daughter and son, that led me into researching this information.

I thank the great professors and other members of the **Conselho Geral do AE de Benfica** that were able to show interest and concern with these issues and were more concerned with how all of this impacts the students, and not themselves.

I thank Professor **Bruno Damásio** at NOVA IMS for the enthusiasm he showed on this topic and for his valuable help, and my dear colleague **Inês Rocha** for helping in discussing and structuring this poster.

Author: Samuel Santos
NOVA IMS Master in Data Science Student (#20220609)
samuel.datasci@gmail.com

Oct-2023

Project:
https://github.com/samueldatasci/ENES
Work in progress; comments are most welcome!