Family, school, and place: Explaining regional differences in student achievement in Portugal before and after COVID-19

Carlos Moreira[1], Carla Sá[2,3,4]

[1] Teach for Portugal, [2] Universidade do Minho, [3] NIPE, [4] CIPES

1. Motivation and Research Questions

Regional disparities in educational outcomes are among the most persistent and policy-relevant dimensions of socio-spatial inequality in European countries. Beyond their social significance, such disparities have long-run implications for regional economic convergence, intergenerational mobility, and human capital accumulation. Where a student grows up and attends school is not merely a geographic fact; it is a powerful determinant of their cognitive development and lifetime opportunities. In countries characterised by pronounced territorial imbalances, the geographic dimension of educational achievement deserves particular analytical attention.

Portugal offers a compelling case study. Despite sustained public investment in education over recent decades and explicit EU cohesion objectives, the country exhibits disparities between coastal and inland regions and between metropolitan and rural areas across a wide range of socioeconomic indicators. These structural asymmetries are expected to be reflected in educational outcomes, yet rigorous microeconometric evidence disaggregated at the regional level remains scarce. Furthermore, the COVID-19 pandemic introduced a sharp exogenous shock to educational systems worldwide. In Portugal, school closures and the abrupt transition to distance learning during 2019-2021 are expected to have amplified pre-existing socioeconomic and spatial inequalities in student learning, but the extent to which this occurred across regions has not been systematically examined.

This paper addresses these gaps by investigating the determinants of academic achievement among Portuguese students aged 15-16, with particular emphasis on regional heterogeneity and its evolution over time. Specifically, it seeks to answer four research questions: (i) What is the role of individual and family characteristics in explaining student performance? (ii) What is the additional role of school characteristics? (iii) Are there significant regional differences in how family and school factors operate? (iv) Did regional disparities in educational achievement widen in the aftermath of the COVID-19 pandemic?

2. Contribution to the Literature

The economics of education literature has long sought to identify the determinants of student achievement through an education production function framework (Todd & Wolpin, 2003; Hanushek, 1997). A broad consensus has emerged that family socioeconomic status (SES) is the dominant predictor of academic performance (Cobb-Clark & Moschion, 2017; Clotfelter et al., 2007), with school-level resources contributing little explanatory power (Lee & Barro, 2001; Krueger, 2003). The role of geographic context, beyond individual and school characteristics, has received comparatively less attention, though a growing body of evidence points to the significance of regional socioeconomic conditions for educational outcomes (Daniele, 2021; Agasisti & Vittadini, 2012; Otero et al., 2021).

For Portugal specifically, Pereira and Reis (2012) provided the first systematic microeconometric analysis of regional differences in student performance using PISA data, documenting substantial variation across Portuguese regions and showing that family background accounts for a larger share of the regional gap than school characteristics. The present paper updates and extends that analysis using PISA 2018 and 2022 data, providing new evidence on the persistence of regional disparities, their structural determinants, and their evolution through the pandemic shock. To our knowledge, this is the first paper to use PISA 2022 data to examine regional educational inequalities in Portugal with a pre/post-pandemic comparative design.

3. Data and Variables

The empirical analysis draws on microdata from the OECD's Programme for International Student Assessment (PISA) for Portugal in 2018 and 2022, complemented by regional-level data from Statistics Portugal (INE). PISA provides internationally comparable test scores in Mathematics and Reading (Portuguese), alongside rich information on student, family, and school characteristics. The 2018 sample covers 5,932 students across 276 schools; the 2022 sample covers 6,793 students across 224 schools.

Given the sampling design of PISA, which does not guarantee sufficient observations at the NUTS III level, we aggregate the 25 NUTS III regions into 11 meaningful geographic units following the methodology of Pereira and Reis (2012): Alto Alentejo, Baixo Alentejo, Vale do Tejo, Algarve, Centro Litoral, Centro Interior, Norte Interior, Norte Litoral, Área Metropolitana do Porto, Área Metropolitana de Lisboa, and Islands. This aggregation preserves sufficient within-region sample sizes for reliable econometric inference while capturing the main axes of territorial heterogeneity in Portugal.

The key outcome variables are student scores in Mathematics and Reading, computed using the ten plausible values provided by PISA. Individual and family controls include gender, grade retention, immigrant status, and the ESCS index (Economic, Social and Cultural Status), which aggregates parental education, occupational status, and home educational resources. School-level variables include school size, class size, teacher quality indicators, school autonomy, availability of ICT resources, extracurricular activities, parental perception of school quality, and several indices of resource and staff shortages constructed from principal questionnaires. Regional controls include GDP per capita, a synthetic regional development index (ISDR), illiteracy rates, pre-schooling rates, retention and dropout rates, the share of population with secondary and tertiary education, physician density, divorce rates, and crime rates.

4. Empirical Strategy

The empirical strategy follows a three-stage approach. In the first stage, we estimate an education production function via OLS with school fixed effects:

$$T_{ijr} = \alpha + \beta F_{ijr} + \gamma \phi_{jr} + \varepsilon_{ijr}$$

where T_{ijr} is the test score of student i in school j and region r; F_{ijr} is a vector of individual and family characteristics; ϕ_{jr} are school fixed effects; and ε_{ijr} is an error term. The inclusion of school fixed effects absorbs all school-level observed and unobserved heterogeneity, allowing identification of individual and family effects. Note that within each region, the school

fixed effects sum to the corresponding regional fixed effect, thereby capturing regional variation net of family composition.

In the second stage, the estimated school fixed effects are regressed on observable school characteristics and regional fixed effects:

$$\hat{\gamma}\phi_{jr} = \eta S_{jr} + \varphi_r + \xi_{jr}$$

where S_{jr} is a vector of observable school characteristics, and φ_r are regional fixed effects. This equation allows us to assess how much of the residual regional variation, after controlling for family and individual characteristics, is explained by school-level factors and how much remains attributable to unobserved regional heterogeneity.

In the third stage, we decompose the variance of student test scores within each region into components attributable to family background, school characteristics, and family-school sorting:

$$Var(T_{ij}) = Var(\beta F_{ij}) + Var(\gamma\phi_j) + Cov(\beta F_{ij}, \gamma\phi_j)$$

The covariance term is of particular interest from a policy perspective, as a positive covariance indicates that students from higher-SES families disproportionately attend better schools, a pattern of sorting that amplifies initial inequalities, while a negative covariance indicates that schools partially compensate for family disadvantage. This decomposition is computed separately for each region and compared across the two PISA waves, allowing for a rigorous assessment of whether the pandemic altered the relative roles of family, school, and sorting in regional inequality.

5. Main Results

The first-stage regression results are consistent across disciplines and years, and broadly in line with the literature. Female students systematically underperform in Mathematics and overperform in Reading relative to their male peers. Grade retention is associated with very large score penalties (80 to 103 points), substantially larger than those associated with other individual characteristics. Immigrant status is negatively associated with performance, though the coefficient becomes statistically insignificant for Reading in 2022. The ESCS index is positively and significantly associated with performance in all specifications, and its coefficient increases between 2018 and 2022, suggesting a growing importance of family socioeconomic background in the post-pandemic period.

Analysing unconditional versus family-conditioned regional gaps relative to the reference region (Área Metropolitana de Lisboa), we find that controlling for family and individual characteristics substantially reduces, but does not eliminate, regional performance gaps. In 2018, the Islands and Norte Interior were the regions most distant from the reference in terms of Reading scores after conditioning. In 2022, regional gaps narrow on average but this convergence reflects a generalised decline in scores across all regions rather than genuine catch-up by lagging regions.

The second-stage results show that school characteristics explain a substantial portion of the residual regional variation in 2018, with the addition of school controls dramatically reducing

the school fixed effect differentials across regions. However, in 2022, schools are markedly less effective at offsetting family-driven disadvantages, with Norte Interior emerging as the most disadvantaged region and the school equalisation effect being considerably weaker.

The variance decomposition reveals several important patterns. First, family background consistently accounts for a larger share of the variance in within-region scores than school characteristics across all regions and years. Second, the explained share of variance declines between 2018 and 2022, indicating greater unobserved heterogeneity in the post-pandemic period. Third, the covariance between family and school components turns negative in some interior and rural regions in 2022 (e.g. Baixo Alentejo and Norte Litoral), suggesting that schools in those regions partially compensate for family disadvantage. Conversely, in Algarve, Centro Litoral, and Norte Interior, the covariance is positive and increases in 2022, indicating that schools are exacerbating pre-existing inequalities, likely reflecting stronger sorting of advantaged families into better schools in those regions.

Regional development indicators — particularly the ISDR and illiteracy rates — are significantly correlated with unconditional performance gaps in 2018, though fewer regional variables retain significance in 2022, possibly reflecting the pandemic shock's dominance in driving cross-regional variation in that year.

6. Conclusions and Policy Implications

This paper provides updated and extended microeconometric evidence on the spatial determinants of educational achievement in Portugal, covering the critical pre- and post-pandemic period. Our results confirm the relevance of family socioeconomic background as a driver of both individual performance and regional inequality, while showing that school characteristics retain meaningful but secondary explanatory power. Territorial heterogeneity is substantial and persistent, embedded in broader patterns of regional socioeconomic development that go well beyond what school policy alone can address.

The pandemic appears to have weakened the equalising capacity of schools, widened regional disparities in some dimensions, and increased the relative weight of family background. These findings carry direct implications for education and cohesion policy. The centralised structure of the Portuguese education system, with nationally uniform curricula and teacher allocation, appears poorly equipped to address the place-specific needs of underperforming regions. Greater school autonomy, place-based resource allocation, and policies targeting family-level disadvantage in lagging regions emerge as priorities. More broadly, the results underscore that educational inequality cannot be addressed in isolation from regional development policy, which is a message of direct relevance to the design of EU Cohesion Policy in the post-2027 programming period.

References

Agasisti, T., & Vittadini, G. (2012). Regional Economic Disparities as Determinants of Students' Achievement in Italy. *Research in Applied Economics*, 4(2), 33-54.

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2007). Teacher credentials and student achievement: Longitudinal analysis with student fixed effects. *Economics of Education Review*, 26(6), 673–682.

Cobb-Clark, D. A., & Moschion, J. (2017). Gender gaps in early educational achievement. *Journal of Population Economics*, 30(4), 1093–1134.

Daniele, V. (2021). Socioeconomic inequality and regional disparities in educational achievement: The role of relative poverty. *Intelligence*, 84, 101-515.

Hanushek, E.A. (1997). Assessing the effects of school resources on student performance. *Educational Evaluation and Policy Analysis*, 19(2), 141–164.

Krueger, A. B. (2003). Economic considerations and class size. *The Economic Journal*, 113(485) F34–F63.

Lee, J.W. & Barro, R.J. (2001). Schooling quality in a cross-section of countries. *Economica*, 68, 465–488.

Otero, G., Carranza, R., & Contreras, D. (2021). Spatial divisions of poverty and wealth: does segregation affect educational achievement? *Socio-Economic Review*, 21(1) 617-641.

Pereira, M.C. & Reis, H. (2012). What accounts for Portuguese regional differences in students' performance? *Banco de Portugal Economic Bulletin*.

Todd, P. & Wolpin, K. (2003). On the specification and estimation of the production function for cognitive achievement. *The Economic Journal*, 113, F3–F33.