# Unraveling the Myths of Rural vs. Urban Academic Achievement Drivers

Ana Beatriz-Afonso [1*], Frederico Cruz-Jesus [1]

[1] NOVA Information Management School (NOVA IMS), Universidade NOVA de Lisboa, Campus de Campolide, 1070–312 Lisboa, Portugal.

## Abstract

The generalized migration of individuals from rural to urban areas is a global phenomenon that entails many divides, education being one of them. However, there is a lack of understanding regarding whether the factors driving higher academic achievement (AA) differ between urban and rural students. This study uses data from almost every student in Portugal who took the Portuguese and/or mathematics high school national exams. By applying OLS, the aim is to identify the AA drivers and compare these drivers between urban and rural areas. Among the key findings, variables related to academic background emerged as the strongest predictors of AA, regardless of the environment. Additionally, ICT access is insignificant in urban and rural areas, while socio-economic status does not significantly impact AA amongst rural students. These findings highlight the need for tailored interventions that address the unique challenges faced by students in different areas, with a particular focus on enhancing academic support structures to improve educational outcomes. To the best of our knowledge, this study is the first to utilize data encompassing virtually every student in an entire country to compare and understand the differences in the determinants of AA between urban and rural areas.

## 1- Introduction

Disparities between rural and urban areas are a global phenomenon marked by significant differences in access to resources and opportunities for life improvement. Numerous researchers have demonstrated that these disparities contribute to pronounced educational inequalities, with students in rural areas facing considerable disadvantages. Rural areas face challenges such as inadequate traditional agriculture practices, a history of population decline due to rural exodus, and lower economic development compared to urban areas. The decreasing importance of agricultural activities and weak productive systems contribute to the overall lower socio-economic status of rural regions [1]. The uneven distribution of resources, infrastructure, and support services has led to a persistent gap in educational outcomes between rural and urban students. Furthermore, educational imbalances persist, as research indicates that the location of schools significantly influences student achievement, with rural and remote schools consistently lagging [2, 3]. This uneven landscape and educational disparities causing struggles for students born in rural areas are shared all over the world, in countries like Germany [4], England [5], Italy [6], Canada [7], Australia [8], United States [9] or even in China [10].

Portugal is no different, exhibiting a strikingly uneven population density distribution [11]. The heterogeneity in Portugal's geography is also seen at the educational level. PISA 2018 results reveal that student performance is strongly asymmetric within the country. An equivalent of two school years separates the country's highest and lowest-achieving regions. PISA also shows that schools have failed consistently to serve as social elevators [12]. Likewise, according to the 2021 Census, the Lisbon metropolitan area comprises 26.6% of its population with tertiary education, significantly contrasting to most rural regions, struggling to reach the 20% mark [13].

---

\* **CONTACT**: aafonso@novaims.unl.pt

Given the heterogeneity in the distribution of resources and opportunities in Portugal, which can be seen as a mirror of other European countries, and previous research proving educational disparities in favor of urban and developed areas, it raises the question of whether a student's region of domicile influences what they need to achieve success in school, i.e., drives academic achievement (AA). Despite existing studies highlighting differences in AA between urban and rural students, the literature remains limited in addressing the distinct needs of the two groups. While research has identified various factors contributing to the achievement gap—such as urban students' reliance on intermediate learning outcomes and personal characteristics versus rural students' dependence on school characteristics [14] and community support [15]—there is a notable gap in directly comparing these influences. The current literature lacks a comprehensive analysis of how these factors affect academic achievement in rural versus urban settings. This gap underscores the need for further research to quantify and understand these differences. Such insights are crucial for informing policy decisions to address regional disparities and ensuring that educational strategies are tailored to the unique needs of students in rural and urban areas. To the best of our knowledge, this is the first study using virtually every student in a compulsory public school system to assess the differences between the drivers of AA in different types of areas. Also, our study uses secondary data collected at an institutional level. This approach helps mitigate the potential biases associated with self-reported data, such as inaccuracies or subjective responses [16]. Our approach employs OLS regression to directly compare the impact of well-established core variables driving AA in both rural and urban student populations. By focusing on national exams, we ensure that all students are evaluated under the same conditions, enhancing the validity of our comparisons. This methodology enables us to pinpoint significant disparities or similarities in how these variables influence AA across different regions, providing robust and actionable insights. In pursuit of this aim, we aim to answer the following research question:

1. *How do AA drivers change between rural and urban areas?*

By addressing this question, we can tailor our support to meet the diverse needs of students, considering the specific characteristics of the regions where they live and study, with solid data evidence. Comparing the needs of rural and urban students contributes to the broader goal of promoting equity in education [14]. Understanding these differences will contribute to the discussion of the competence of the educational system, namely the efficiency of a centralized education system and the autonomy of schools [17].

In answering the research question, this study is structured as follows: Section 2 presents the literature review; Section 3 describes the methods used; the results are presented in Section 4, whereas Section 5 discusses the results and implications. Finally, the conclusions are outlined in Section Six.

## 2- Theoretical Background and Hypotheses

AA is defined as the outcomes and abilities of students across various school subjects contributing to their success [18]. It serves as a pivotal factor in shaping life trajectories, influencing opportunities such as access to higher education, employment prospects, and self-perceptions [19]. Given the broad nature of academic achievement, numerous variables can be employed to measure it. In this study, we adopted the transition of the academic year, as suggested by Cruz-Jesus et al. [20].

As previously mentioned, the presence of AA disparities between rural, often less developed areas and urban areas has been consistently demonstrated by various authors across the globe. Notably, rural areas frequently show a higher percentage of underperforming students than their urban counterparts [8]. It is imperative to assess the variations among the primary factors influencing AA to comprehend the distinct requirements for student success.

**Academic Background**

Numerous studies have consistently demonstrated that a student's prior AA serves as the most robust predictor of their current and future academic performance [21-23]. Remarkably, this fundamental reality persists regardless of whether the educational context is rural or urban [3]. Moreover, Gordon & Monastiriotis [5] observed that urban areas typically exhibit more significant variability in exam results, with particular schools excelling while others falter. In contrast, rural areas tend to display less disparity in exam outcomes, with most schools performing at similar levels. These disparities may stem from disparities in resource accessibility, student population stability, demographic shifts, and community characteristics. These findings lead us to the following hypotheses:

*H1: Previous AA is positively associated with AA.*

Regardless of the educational setting, academic struggles leading to grade repetition can significantly delay long-term academic attainment [24]. Despite educators' initial belief in the beneficial effects of grade retention, research indicates that these benefits diminish over time, particularly in later grades [25, 26]. Furthermore, grade retention usually leads to negative consequences, including peer stigmatization, reduced self-esteem, and diminished commitment to schooling [27]. Considering the impact of previous achievement and retention, we consider the following hypotheses:

*H2: Previous retention is negatively associated with AA.*

**ICT Access**

The proven differences in the characteristics of the areas will also lead to different opportunities for students in more urban and developed areas compared to those in more rural ones. In rural areas, students might face challenges due to limited access to educational resources and opportunities, affecting their academic success [28, 29]. On the other hand, urban areas usually offer better access to resources like well-equipped schools and extracurricular activities than rural areas [4]. Unlike their rural peers, urban students may have more chances for academic growth with exposure to activities like internships and research opportunities [3]. Even at a very early stage, children from a more diverse and cultural environment tend to have better AA [30].

Considering the differences found among the two types of students, it is essential to analyze the influence of one of the most debated drivers of AA today – the use of Information and Communication Technology (ICT). While differences in computer ownership have become less significant [29], students with less access to computers or other ICT resources, which happens mainly in rural areas, usually perform lower than their peers [15]. We therefore propose:

**H3:** *ICT access is positively associated with AA.*

**Gender**

The role of gender in AA has remained a contentious subject in scholarly discourse for several decades. Some studies suggested that gender was not a significant factor in AA [31, 32]. However, others suggest that gender indeed exerts a tangible influence on academic outcomes [21]. Valli Jayanthi et al.'s [33] study has indicated that females often exhibit higher academic performance. Musso et al. [34] introduce a different perspective, attributing this disparity to inherent gender-specific characteristics and their varied responses to environmental stimuli. For this reason, it is crucial to understand if the gender impact exists and if it differs in urban versus rural areas. Consequently, we propose the following hypotheses:

**H4:** *Female gender is positively associated with AA.*

**Socio-Economic Status**

Examining regional disparities, it becomes apparent that resource inequalities, socio-economic status (SES), and family background variations are significant differences among students from different areas [15]. These differences lead us to a major challenge, as many authors emphasize the considerable role of SES as a critical driver of AA. These findings are particularly relevant when examining the diverse economic landscapes across regions, as seen in Portugal's contrasting dynamics among urbanized, metropolitan areas and more rural interior regions, which experience varying levels of economic development and industrialization [2]. The SES of students emerges as a crucial factor influencing their AA, especially for those underachieving or residing in remote locations [4]. Due to the characteristics of the population in urban areas, schools in more affluent areas tend to perform better than those in less affluent areas.

Nevertheless, urban areas tended to have more social segregation in terms of school choice, with families from higher socio-economic backgrounds more likely to travel further to attend a school perceived as better [6]. Although SES is a widely debated factor of AA, there are several different ways of measuring it, such as parental education, parental occupation, and income [35]. In Portugal, families with lower incomes have the option to apply for state-granted educational allowances to alleviate the financial burden of education. Research consistently shows that lower SES correlates with poorer academic performance [25]. Therefore, using educational allowances as a proxy for lower-income families, we propose:

**H5:** *Students receiving social support is negatively associated with AA.*

Parental education serves as a crucial SES proxy and a pivotal determinant of academic success for children. Research by Steinmayr et al. [36] highlights the significant influence of parental education on children's motivation and educational achievement. Moreover, parental education reflects the value of education within the household, with parents possessing lower educational backgrounds often struggling to prioritize and support their children's academic endeavors [37, 38]. This disparity in parental education becomes particularly salient in understanding the rural-urban educational gap [11]. Generally, parents with higher levels of education are more likely to secure better-paying jobs, leading to enhanced socio-economic status and often residing in urban areas where educational opportunities abound.

Consequently, these parents are more inclined to prioritize and invest in their children's education, translating into improved academic performance [39]. This phenomenon can be supported by the weak productive system in rural areas, which may result in fewer economic opportunities compared to urban areas [2]. Given the importance of the parents' education in several aspects of the student's life, we propose:

**H6:** *Legal guardian's education is positively associated with AA.*

### 2-1- Conceptual Model

Figure 1 presents our proposed model based on previous AA literature. For each construct, three research hypotheses were developed, one relating to general AA and the other two derived from the urban and rural environments, thus allowing us to compare the impact of each determinant in these areas.
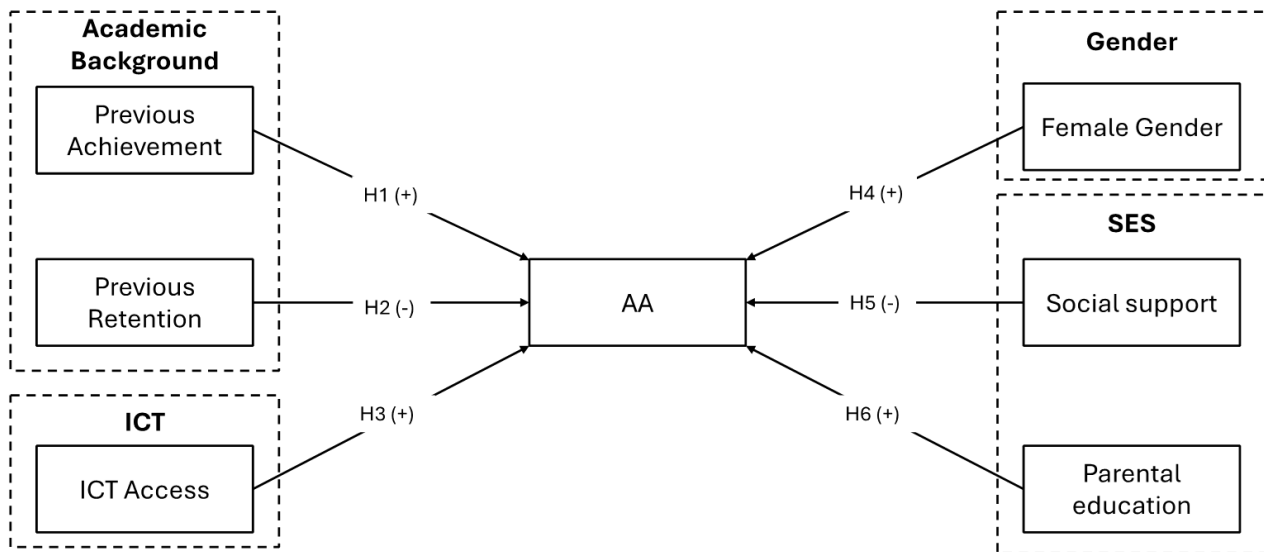


**Figure 1. Conceptual Model and Hypotheses**

## 3- Data and Methodology

### 3-1- Data

This study utilized an anonymized dataset from the Directorate-General of Statistics for Education and Science (DGEEC) of the Portuguese Ministry of Education, encompassing information about nearly all students in the compulsory schooling system. Data from the academic year 2021/2022 was employed, representing the latest available information, as the exams always present a year delay. After data treatment, the dataset included details on 38,754 high school students, with 25,420 attending schools in urban areas and 13,334 in rural areas, all of whom took at least one of the national exams in Portuguese or mathematics.

Since the 2020/2021 academic year, national exams in Portugal have undergone significant changes. These exams now exclusively serve as entry exams for tertiary education, meaning only students aspiring to pursue higher education typically sit for them. Additionally, students have the autonomy to select which exams to take based on their desired tertiary degree requirements. The national exams comprise two phases. The first phase permits students to utilize the exam results in any of the three stages of the National Competition for Access to Higher Education. Conversely, the second phase restricts the use of exam results to the latter two stages of the competition.

Portuguese and mathematics are the most selected among the available exams, as they align with the entry requirements of numerous university programs across various fields of study. For students who undertake or repeat both exams, the target variable comprises an average of their grades obtained in one or both exams across one or both phases.

To refine the analysis, an initial preselection process was implemented to ensure students were in similar conditions. As such, it focuses exclusively on students pursuing regular studies, excluding those in professional and artistic programs. Additionally, the study only uses students enrolled in public schools due to the substantial differences between them and private schools [40].

### 3-2- Methodology

#### Data Preprocessing

From a contextual perspective, it is essential to acknowledge that schools use two different databases for data submission. The absence of a standardized format and the asynchronous nature of data submission result in missing values, errors, inaccuracies, and duplicates in the dataset. We excluded all variables with over 60% missing data and all students who did not present the previous grades. After this process, we still had some variables with a maximum of 20% missing data, leading to the need to employ an imputation method, in this case, K-Nearest Neighbors (KNN). KNN functions by identifying the K nearest neighbors of an instance with missing data and using their information to estimate the missing values. The similarity between instances is measured through a distance function, such as the Euclidean

distance. The missing value is then estimated using the mean or median of the values of the nearest neighbors. KNN stands out as a non-parametric algorithm, making no assumptions about the data distribution. It has proven to be a robust tool for accurately handling substantial sets of missing data [41].

Following the cleaning process to solve the mentioned issues, one-hot encoding is applied to represent categorical variables as binary vectors. The one-hot encoding procedure involves mapping categorical values to integer values and subsequently converting each integer value into a binary vector. In this vector, all values are zero except for the index corresponding to the integer, which contains a one. This transformation becomes necessary when there is no ordinal relationship between the categories, mitigating biases associated with integer representations [21]. To avoid such type of feature bias, quantitatively, the features were transformed into the same scale where each feature had zero mean and unit variation [42].

Additionally, as in Bilal et al. [41], decision trees were used to define decision rules. In this case, the rules were the breakeven point where a school and a class start to be considered significant. A school's location was classified as rural based on its affiliation with a municipality predominantly characterized by rural parishes. To determine these municipalities, we referred to the list of rural parishes outlined in the "Programa de Desenvolvimento Rural 2014-2020" [42].

### *Ordinary Least Squares*

Initially, the dataset was divided into three subsets: train, validation, and test. The train and validation sets were utilized for model training and preventing overfitting. In contrast, the test set was reserved for evaluating model performance and assessing the influence of independent variables on the dependent variable.

Ordinary Least Squares (OLS) is a specific type of regression model and a robust tool for examining relationships [8]. Among OLS's several advantages, we can highlight that it provides unbiased estimates of the coefficients of the independent variables, ensuring that the estimated coefficients, on average, equal the actual population values. This feature enhances the reliability of OLS in estimating relationships between variables. Also, OLS produces efficient estimates by minimizing the variance of the estimates. This minimization results in more precise and reliable estimates. OLS is also relatively easy to interpret, as the coefficients signify the change in the dependent variable for a one-unit change in the independent variable while holding all other variables constant. Lastly, OLS is a flexible method applicable to a broad range of research questions [43]. Moreover, OLS regression is widely used for estimating relationships in linear regression models. It allows researchers to analyze the impact of independent variables on the dependent variable, assess the significance and strength of these relationships, and control for other relevant factors [44].

OLS takes the following form [45]:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p + \epsilon \tag{1}$$

Where $X_p$ represents the $jth$ predictor, $\beta_p$ the slope of the coefficient for each explanatory variable, $Y$ the dependent variable, and $\epsilon$ the residuals.

OLS was chosen as the primary analytical method for this study due to its well-established effectiveness in exploring relationships between variables, mainly when the goal is to interpret and understand the underlying factors influencing AA. While more advanced machine learning techniques may offer superior predictive accuracy, they often suffer from the "black-box" effect, where the model's inner workings are not easily interpretable. Our primary objective was not to predict AA with the highest accuracy but to gain a deeper understanding of the factors driving it. OLS has a long history of successful application in various studies examining AA and related factors. Among many other utilizations, it has been used to analyze the impact of student background variables [46], sociodemographic factors [7], peer influences [47], identify potential dropouts [48], and mediate the role of cultural influences [49]. These studies demonstrate the robustness and versatility of OLS in similar contexts, making it a suitable choice for our analysis. Other statistical models were considered, but they were ultimately not employed due to their limitations in interpretability, which is central to our research objectives. OLS's transparency allows for a more straightforward interpretation of the results, providing valuable insights that align with our study's primary aim.

## 4- Results

### *4-1-Independent Variables*

Table 1 outlines the hypothesized explanatory variables to facilitate the conceptual model's empirical testing. However, it is essential to note that due to limitations inherent in the dataset, there is a possibility that other drivers of AA are not being captured.

**Table 1. AA drivers and their respective support**

| Code | Variable | Support |
|---|---|---|
| St_Prev_AA | Student GPA in the year before performing the exam. | Miguéis et al. (2018), Coleman (1968), Ferrão & Almeida (2019), and Barbosa-Camacho et al. (2022) [23, 37, 50, 51] |
| St_Age_Dif | Student age difference against the average of the class. Proxy for retention. | Cruz-Jesus et al. (2020), Guèvremont et al. (2007) and Nunes et al. (2022) [20, 52, 53] |
| St_PC | The student has a computer at home (Y=1/N=0). | Lee (2022), Costa-Mendes et al. (2020) and Kubey et al. (2001) [14, 32, 54] |
| St_Fem | The student is a female (Y=1/N=0). | Cruz-Jesus et al. (2020), Jayanthi et al. (2014), Musso et al. (2020) and Mensah & Kiernan (2010) [20, 33, 34, 55] |
| St_SS | The student has social support (Y=1/N=0). | Archibald (2006) and Delen (2010) [30, 56] |
| EE_Educ_Sup | EE has tertiary education (Y=1/N=0). | Nunes et al. (2022) and Chesters & Daly (2017) [53, 57] |

To test the conceptual model, we created one model for all students and two additional models: one for students attending schools in urban areas and another for students attending schools in rural areas. These OLS models can be represented mathematically as:

$$AA_i = \beta_0 + \beta_1 \times St\_Prev\_AA_i + \beta_2 \times St\_Age\_Dif_i + \beta_3 \times St\_PC_i + \beta_4 \times St\_Fem_i \\ + \beta_5 \times St\_SS_i + \beta_6 \times LG\_Educ\_Sup_i + \varepsilon_i \qquad (2)$$

where $\beta_0$ is the constant term, and $\beta_1$ to $\beta_6$ are the coefficients to be estimated by the OLS, $\varepsilon_i$ is the error term of the $i^{th}$ student. St_Prev_AA = the GPA of the $i^{th}$ student in the year before, St_Age_Dif = difference between the age of the $i^{th}$ student and their class, St_PC = 1 if the $i^{th}$ student has a computer at home, St_Fem = 1 if the $i^{th}$ student a female, St_SS = 1 if the $i^{th}$ student has social support, and LG_Educ_Sup = 1 if the $i^{th}$ student legal guardian has tertiary education.

### *4-2- OLS Results*

The first step was to analyze Pearson's linear correlation coefficients among all variables, as shown in Table 2. All the variables have a statistically significant correlation ($p<0.001$), with none of the independent variables being highly correlated.

**Table 2. Correlation Matrix**

| | St_Prev_AA | St_Age_Dif | St_ PC | St_Fem | St_SS | LG_Educ_Sup | AA |
|---|---|---|---|---|---|---|---|
| **St_Prev_AA** | 1 | | | | | | |
| **St_Age_Dif** | -0.23*** | 1 | | | | | |
| **St_ PC** | 0.07*** | -0.5*** | 1 | | | | |
| **St_Fem** | 0.08*** | -0.04*** | -0.02*** | 1 | | | |
| **St_SS** | -0.03*** | 0.02*** | -0.18*** | 0.05*** | 1 | | |
| **LG_Educ_Sup** | 0.18*** | -0.09*** | 0.11*** | -0.06*** | -0.23*** | 1 | |
| **AA** | 0.65*** | -0.19*** | 0.05*** | 0.08*** | -0.06*** | 0.2*** | 1 |

Table 3 shows the result of the OLS models (global, urban, and rural). The authors conducted a set of tests to confirm the suitability of OLS to our data. The Durbin-Watson test yielded results around 2, indicating no issues with first-order autocorrelation in the models [58]. Following the recommendations of Anscombe & Tukey [59], who advocate for the graphical representation of residuals due to their higher sensitivity, we employed scatter plots to evaluate each variable and histograms to assess the distribution of the residuals. These distributions and QQ plots are provided below in Figure 2-4.

The histogram plot illustrates the differences between observed and predicted values (residuals) and assesses how well these residuals align with a normal distribution. Across all three models, the residuals closely follow the normal distribution curve, with some minor deviations. The Q-Q plots further evaluate the residuals' normality by comparing the theoretical quantiles (x-axis) with the actual residuals (y-axis) quantiles. Residuals are considered normally distributed when the points align with the red line. Although the alignment is not perfect in all three cases, most points fall close to the line, indicating that the residuals approximate a normal distribution.

Thus, we can reasonably conclude that the residuals follow a normal distribution upon analyzing these visualizations. Variance Inflation Factors (VIF) were calculated to assess the degree of multicollinearity among the independent variables [60]. Our analysis indicates no substantial collinearity, as all VIF values are below 10 [60, 61]. The correlation matrix presented earlier also supports the absence of multicollinearity. To test for heteroscedasticity, we applied White's test [62], which confirmed the absence of heteroscedasticity in our data ($p>0.10$).
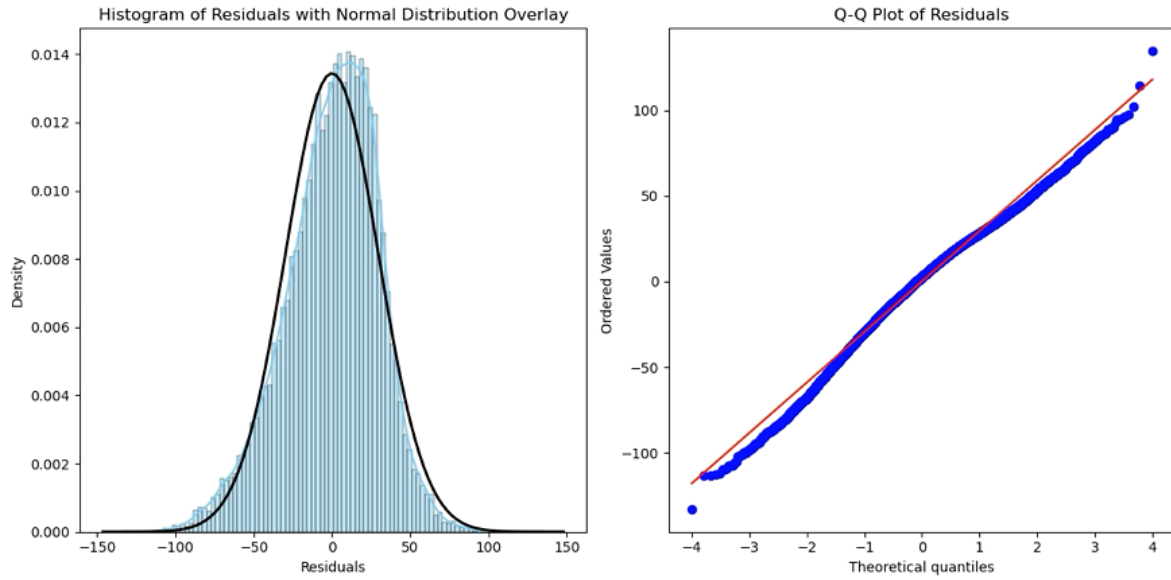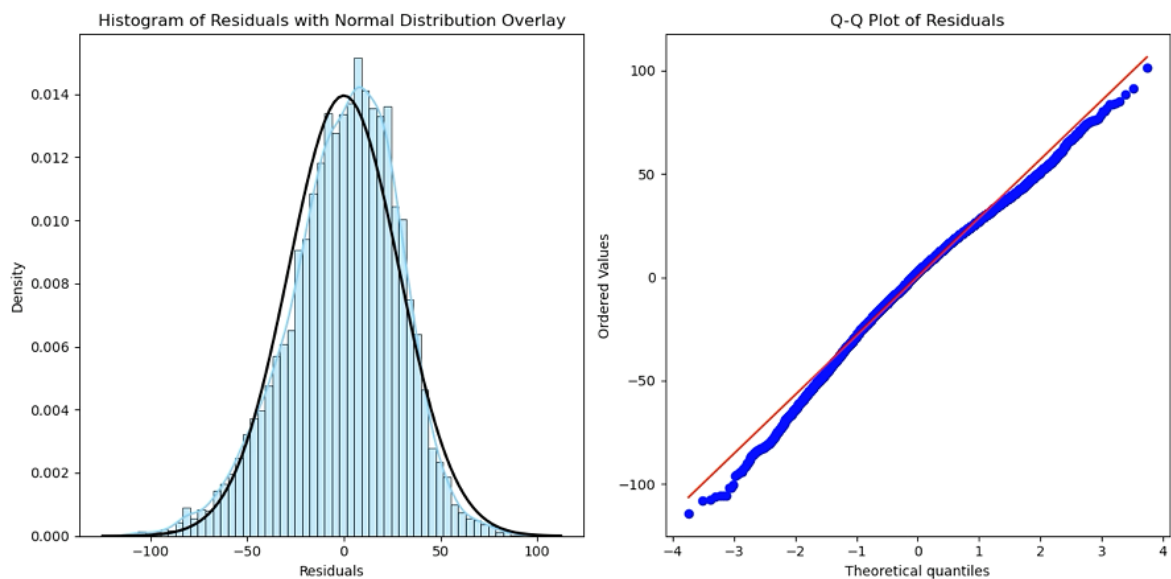
**Figure 2.** *Global model – Residuals' distribution and Q-Q Plot*
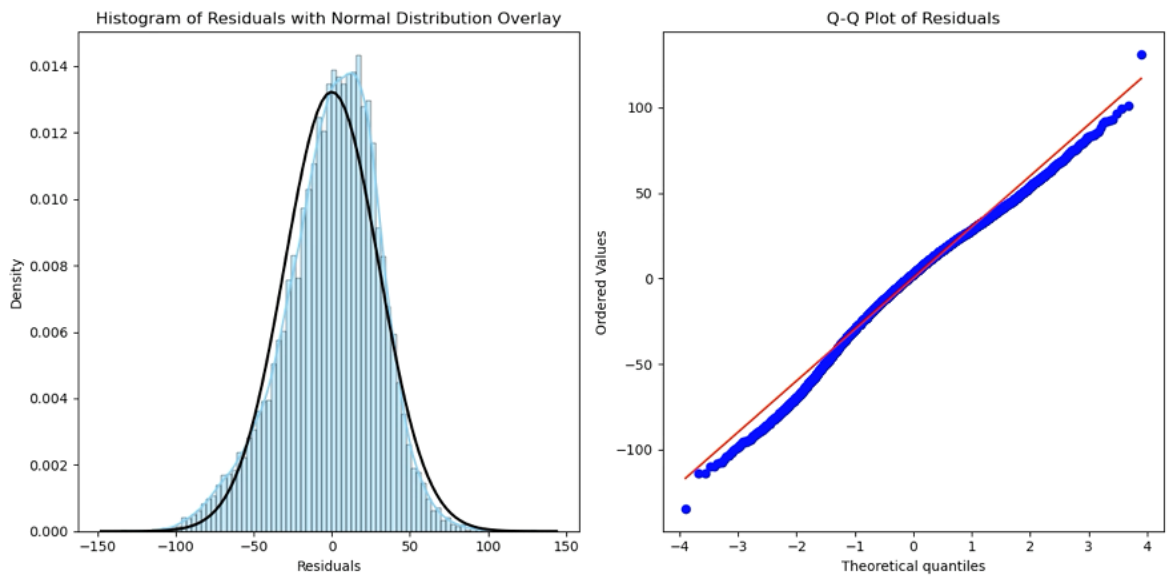


**Figure 3.** *Urban model – Residuals' distribution and Q-Q Plot*



**Figure 4.** *Rural model – Residuals' distribution and Q-Q Plot*

**Table 3. OLS results for Global, Urban, and Rural models**

| | Global | | Urban | | Rural | |
|---|---|---|---|---|---|---|
| | $\hat{\beta}$ | VIF | $\hat{\beta}$ | VIF | $\hat{\beta}$ | VIF |
| **Intercept** | 126.221*** | | 125.792*** | | 126.707*** | |
| **St_Prev_AA** | 24.650*** | 5.792 | 24.422*** | 5.892 | 24.903*** | 5.736 |
| **St_Age_Dif** | -1.229*** | 1.069 | -1.306*** | 1.053 | -1.483*** | 1.079 |
| **St_PC** | 0.092 | 3.216 | -0.217 | 3.372 | 0.154 | 3.208 |
| **St_Fem** | 1.433*** | 2.250 | 1.616*** | 2.332 | 1.287*** | 2.215 |
| **St_SS** | -0.918*** | 1.370 | -1.205*** | 1.484 | -0.732 | 3.208 |
| **LG_Educ_Sup** | 2.987*** | 1.731 | 2.726*** | 1.545 | 2.573*** | 1.822 |
| **N** | 21702 | | 14235 | | 7466 | |
| **R-Square** | 0.432 | | 0.420 | | 0.452 | |
| **F-Statistic** | 2755*** | | 1720*** | | 1026*** | |
| **Durbin-Watson** | 1.989 | | 1.996 | | 2.008 | |

Analyzing the global model, we observe an R-squared value of 0.432. This result indicates that with just seven variables, the model successfully explains over 43% of the variance in AA across a dataset of 21,702 students. This level of explanatory power is significant, especially considering that these variables are derived from secondary data— information that all relevant school professionals and policymakers have access to. Regarding the hypotheses, except for St_PC (H3: $\hat{\beta}_3 = 0.092$, p-value > 0.10), which contrasts previous studies [33], all the other variables are confirmed to be statistically significant drivers of AA. Specifically, St_Prev_AA (H1: $\hat{\beta}_1 = 24.650$, p-value < 0.01) is consistent with the literature as the strongest predictor of AA [24]. St_Age_Dif (H2: $\hat{\beta}_2 = -1.229$, p-value < 0.01) serves as a proxy for retention and is also recognized as a clear driver of AA by other authors [21]. St_Fem (H4: $\hat{\beta}_4 = 1.433$, p-value < 0.01) aligns with recent literature suggesting that females tend to outperform males [63]. Both proxies for SES have confirmed their relevance: St_SS (H5: $\hat{\beta}_5 = -0.918$, p-value < 0.01), which aligns with previous literature [25, 31], and LG_Educ_Sup (H6: $\hat{\beta}_6 = 2.987$, p-value < 0.01), recognized as one of the strongest drivers of AA in earlier studies [39].

In the models divided by location, the urban model accounts for 14,235 students and explains 42% of the variance. A similar scenario of the global model is observed for urban students, with St_PC (H3a: $\hat{\beta}_3 = -0.217$, p-value > 0.10), which does not support the positive influence of ICT on AA in urban students. This finding contrasts with findings from Lee [14] on PISA 2018 results from China, where ICT use was significant for urban and rural students. However, the remaining factors are consistent with the literature and the global model, including St_Prev_AA (H1: $\hat{\beta}_1 = 24.422$, p-value < 0.01), St_Age_Dif (H2: $\hat{\beta}_2 = -1.306$, p-value < 0.01), St_Fem (H4: $\hat{\beta}_4 = 1.616$, p-value < 0.01), St_SS (H5: Beta = -1.205, p-value < 0.01), and LG_Educ_Sup (H6: $\hat{\beta}_6 = 2.726$, p-value < 0.01) are supported, in line with the literature as mentioned in the global model. A detailed discussion of these findings and their implications is provided in the next chapter.

In the rural model, which includes 7,466 students and explains 45.2% of the variance, the positive impact of ICT on AA St_PC (H3: $\hat{\beta}_3 = 0.154$, p-value > 0.10) and the negative impact of social support on AA St_SS (H5: $\hat{\beta}_5 = -0.732$, p-value > 0.10) are not supported, diverging from some of the existing literature [3, 63]. Nonetheless, consistent with the global and urban models, St_Prev_AA (H1: $\hat{\beta}_1 = 24.903$, p-value < 0.01), St_Age_Dif (H2: $\hat{\beta}_2 = -1.483$, p-value < 0.01), St_Fem (H4: $\hat{\beta}_4 = 1.287$, p-value < 0.01), and LG_Educ_Sup (H6: $\hat{\beta}_6 = 2.573$, p-value < 0.01) appear to be supported.

## 5- Discussion

### 5-1- Discussion of Findings

In this study, we compared the contribution of several key factors driving AA, as identified in previous literature, to understand what drives AA and the differences between AA drivers in urban and rural areas. We used virtually every student in Portugal who performed the high school Portuguese and/or mathematics national exam. Figure 5 shows the details on the proposed models and the results of this research.
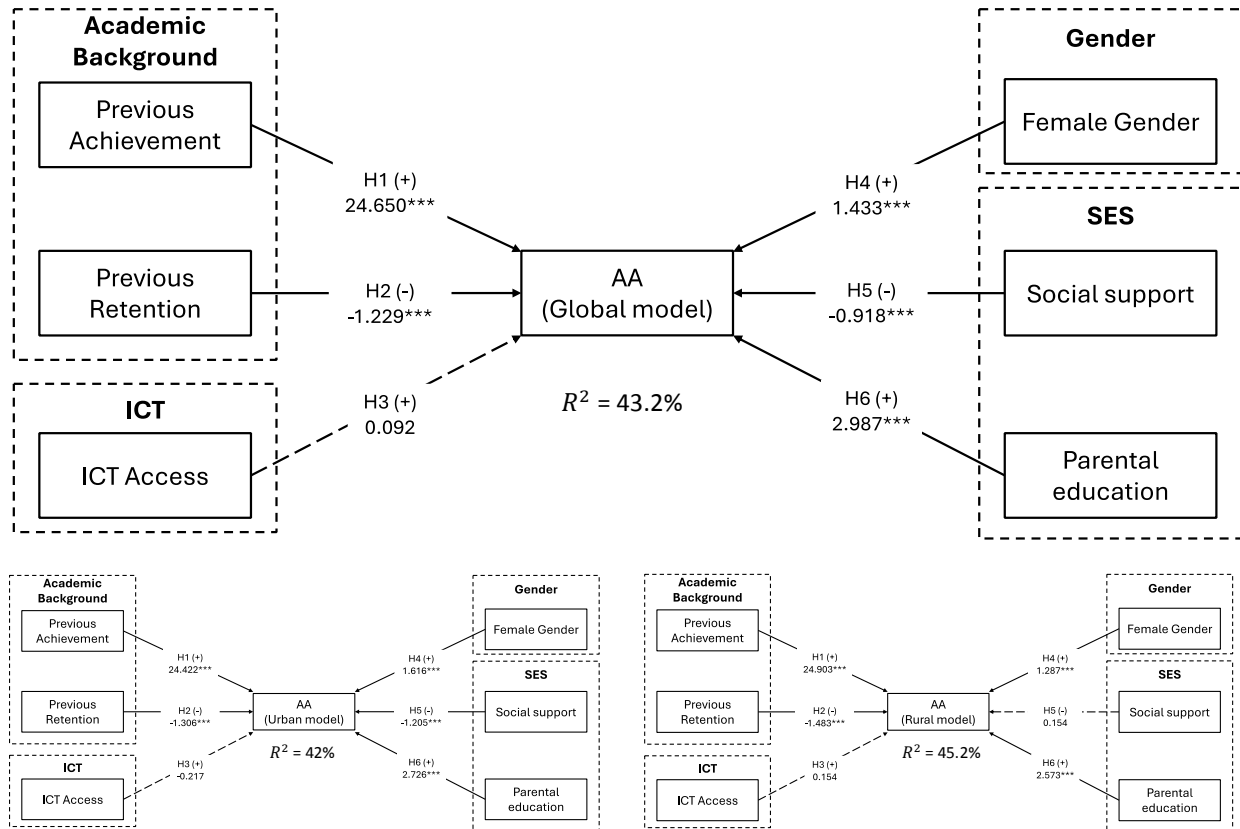
**Figure 5.** *Structural model results: global model (upper-half); urban model (lower-left); and rural model (lower-right)*

It was no surprise that previous GPA emerged as the most significant variable across all three models, consistent with the findings of previous literature. As a proxy for retention, the age difference compared to the class average also aligned with our expectations. Students with a more significant age difference from their peers—those most likely to have repeated a year—tended to have lower AA. Thus, the hypotheses regarding academic background are all confirmed, no matter the environment, in coherence with previous literature [4].

The rationale for retention is that repeating a grade allows students to master the material needed to succeed in subsequent grades [64]. Although some studies indicate that retention can lead to minor improvements in AA in the early grades [65], this effect tends to diminish in older students, particularly in high school [66]. Other research has shown that retention can not only harm AA but also negatively impact academic self-concept, homework completion, and self-esteem while promoting maladaptive motivation and increased absenteeism [67]. In the long term, retention can have significant negative consequences on students' academic choices and future earnings [68]. Being retained is a strong predictor of dropping out, as students who are held back may experience feelings of failure and struggle to catch up with their peers, leading to disengagement from school [48]. In Portugal, students with less maturity and worse economic conditions are more likely to fail a grade, thereby perpetuating their status. Thus, given the high costs and limited benefits of retaining students for an extra year, reevaluating grade retention policies is important [65]. Policies that promote intervention and robust support systems are essential, including professional development for teachers and school staff to better assist struggling students [64]. Also, preventive measures are vital to ensure students receive support before retention becomes necessary [67]. Monitoring and evaluating the effectiveness of retention practices can help identify the most successful strategies, allowing for more efficient resource allocation. By focusing on preventive measures, schools can lead to better educational outcomes and reduce the negative impacts of repeating a grade [68].

Freeman & Simonsen [69] proposed a systematic review of the impact of policies and interventions on high school dropouts, concluding that the most effective approaches are multicomponent interventions, early intervention, and strategies addressing the school organizational structure. Among successful policies to prevent retention and early dropout in Europe, Finland's Flexible Basic Education (FBE) is notable. Targeting middle school students at risk of dropping out, FBE starts in years 7-9 and is managed by local authorities. Students in regular classes who face challenges can join FBE after demonstrating motivation and receiving a special education decision. This program offers flexible learning options within the national curriculum, such as activity-based learning and small-group instruction [70]. It has shown success, with 90% of students improving performance and reducing dropout rates. In Belgium, modular education addresses early dropout by dividing conventional courses into smaller modules, each leading to partial certificates that accumulate into a full qualification. This approach has proven effective in reducing dropout rates [71]. In Portugal, modular education is available only for students over 18, typically through evening classes. Extending this method to

younger at-risk students in traditional high school settings could be beneficial. It would provide flexibility, partial certifications, and regular feedback, boosting students' confidence, motivation, and commitment by helping them set and achieve short-term goals [71].

Unexpectedly, in the era of information technology and contrary to most research findings, our results indicated that ICT access was not a significant AA driver across all three models. This aspect may suggest that owning a computer is no longer a distinguishing factor. In 2022, 83.9% of the Portuguese population owned a personal computer [72], and 89% had internet access and usage [73]. In particular, after COVID-19, the number of students with no access to ICT has significantly decreased, largely due to an action part of the "Recovery and Resilience Plan" program that provided computers and hotspots to over 600,000 students [74]. Instead, factors such as ICT literacy [75], competence, type of use, or even interest [76] may be more relevant in understanding the differences in academic achievement. Additionally, Park & Weng [76] emphasize that these attributes in the usage of ICT are often lacking in low-income families. Ren et al. [77] further revealed that factors such as cultural resources, cultural practices, and active parental mediation are significant predictors of adolescents' digital skills, creative skills, and educational use of the Internet. As previously discussed, these factors can differ significantly between urban and rural areas. The possibility of having an inquiry into the type of usage and test for the ICT skills of the students, could be strongly beneficial for policy makers to know where to act.

In both cases, in urban and rural areas, females significantly perform better than males, which goes hand in hand with previous literature [32, 33, 78]. This phenomenon is not related to gender alone but more to inherent characteristics generally associated with one of the genders. Factors such as greater motivation, a more positive attitude toward school, higher goals at different levels, and a more adaptive profile are characteristics prone to females [79]. A study performed by Verbree et al. [78] showed that conscientiousness, i.e., being careful, responsible, organized, and achievement-oriented, among others, found conscientiousness fully mediated the gender gap in achievement, even when controlling for prior achievement.

The SES variables are among the most important for the students in general. It is vital to notice that rural students are more likely to come from a low SES background and have parents with lower levels of education. In this study, we use the most used national exams to enter university, Portuguese, and mathematics. Nevertheless, even after they have completed these exams, students from rural areas face greater challenges when pursuing higher studies. The costs associated with attending university, including relocation expenses, are higher for regional and rural students. Financial support from the government may not cover these costs, making it difficult for students from low SES families to relocate to university. Although students from rural areas face greater challenges, it is interesting that the negative impact of needing social support for students is not evident in rural environments, contrary to the global and urban model. This finding aligns with previous literature indicating a weaker relationship between family SES and academic achievement in rural areas, as more factors are intervening [63]. Wang's [49] study suggests that monetary influence is less critical for rural students' achievements, which rely more on cultural factors. Cultural capital, such as books, art, and educational material, encompassing the cultural knowledge and skills valued in the educational system and often transmitted within families, plays a crucial role. It positively correlates with academic achievement and significantly mediates the relationship between family SES and AA [80].

Although cultural capital is not directly evaluated in our data, these cultural capital factors are usually closely linked to parental cultural capital, their education, and their involvement in their children's education [81]. In line with previous findings, our study shows that having legal guardians with tertiary education is significant and has a high coefficient for all three models. Young people with highly educated parents are more likely than their peers with less educated parents to attain high school achievement levels. Specifically, students with university-educated parents achieve outcomes at much higher levels than their peers with low-educated parents [57]. Educated parents tend to have and communicate higher expectations for their children [82], motivating the students to have greater self-regulation and higher educational aspirations, promoting AA. Not only at home, but a school's "collective expectation" is also paramount. Collective expectations may be more pronounced in urban schools due to higher levels of segregation or more effective in rural schools due to stronger social bonds [83]. It is essential to recognize that other cultural influences may also impact students. Among others, community influence, particularly perceptions of job opportunities, can negatively affect educational aspirations in rural areas. Agger et al. [84] found that educational aspirations fully mediated the relationship between perceptions of job opportunities in the rural community and university enrollment. This finding suggests that adolescents' educational goals are shaped by their perceptions of family and community, translating into actual enrollment behaviors. However, stronger community support in rural areas often positively drives educational aspirations [16]. One crucial non-observed factor is the regional variance in the value attributed to knowledge, which shapes the students' self-expectations and motivation. To reduce this gap, rural communities and families need to continue supporting their children and enhancing their self-esteem and motivation for higher AA.

In rural areas, creating an environment that promotes what these areas lack to achieve higher levels of AA is crucial. This includes providing extracurricular activities, museums, and better school equipment, among other resources. Additionally, parental involvement should be encouraged, and it is essential to ensure that teachers are culturally aware

of the unique needs of these communities. Increasing instructional time for key subjects, particularly for students from lower SES backgrounds, is also vital [85]. Regarding post-secondary education, another important aspect is the provision of greater state assistance for students who need to relocate from rural to urban areas to pursue higher studies. This support would motivate students to strive for admission to top universities. Without policies to mitigate the exodus and to give the same opportunities to the rural population for school and work, what will continue to happen is that students who move to urban areas for higher education do not return as they do not see opportunities, leading to a loss of talented youth and continuing the desertification of the rural areas [86].

Although the study compares urban and rural areas, greater heterogeneity of results is found within urban areas. Urban areas tend to exhibit higher levels of segregation, encompassing both the best and the worst schools [6]. This segregation is closely related to residential segregation and rigid school catchment area policies based on students' place of residence. This factor contributes to disparities in the overall SES of schools, the individuals attending them, and the opportunities provided [87]. Students are influenced by their peers' abilities [88] and the school environment, including factors like drugs, violence, and resources [89]. This byproduct of segregation leads to social inequality, reduced social cohesion, and limited upward mobility [90], resulting in the creation of "elite schools" with high levels of ability and overall higher SES, which adversely affects low SES students who are most impacted by school characteristics [91].

To address this phenomenon, it is imperative to provide additional resources to disadvantaged schools, enforce and possibly redefine school catchment areas to promote social mixing, and regulate the real estate market to reduce socio-economic and ethnic segregation within urban spaces [87]. Successful policies, such as temporary funding and partial debt relief in Hessen, Germany, offer valuable lessons for promoting equality [92]. These strategies could also be effective in Portugal, including stable public budgets, democratized decision-making, and community-driven projects. Portugal's Program Educational Territories of Priority Intervention (TEIP), established in 1996, has significantly reduced school failure and dropout rates in disadvantaged areas. However, gaps persist in both internal and external evaluations, such as school grades and national exam results [93]. To further address these challenges, limiting school choice options, exacerbating segregation, ensuring equitable funding for public schools, and promoting diverse and inclusive school environments are crucial steps [90]. Progressive ruralism is crucial when addressing educational disparities between students from different areas. This approach seeks to enhance rural education by integrating interdisciplinary methods, empowering local communities, and addressing social justice issues, all contributing to sustainable development. These strategies have been successfully implemented in various contexts in the United States, promoting equity and sustainability in rural education [94].

### 5-2- Limitations

Some limitations need to be addressed. Firstly, as this study relies solely on secondary data, there is a constraint in the range of available variables. Integrating variables related to personality traits, mental health, and ICT usage could enhance the breadth of analysis, offering a more comprehensive understanding of the distinctions among students in rural and urban areas. The considerable disparity in student numbers between urban and rural locales warrants attention. In the future, researchers should explore methods for sampling balance to mitigate this imbalance effectively. Finally, it is important to note that this study exclusively focuses on students who opt to take exams, primarily motivated by their desire to pursue higher education. As a result, the data may not fully reflect the performance or characteristics of students who, for various reasons, do not pursue university or may have left the educational system before reaching this stage. Consequently, the exam results and subsequent analyses might not account for the broader diversity of student experiences and abilities, potentially skewing the interpretation of how national exams correlate with academic outcomes and higher education aspirations. Future research should broaden its scope to encompass various educational cycles, providing a more comprehensive understanding of the topic.

## 6- Conclusion

Despite existing research highlighting differences in AA between urban and rural students, there is a notable gap in studies that directly compare these influences across environments. Our study addresses this gap using a robust dataset that includes virtually every student in Portugal's public school system. This allows for the direct comparison of AA drivers in urban and rural areas. By employing OLS regression and focusing on standardized national exams, we ensure that our analysis is rigorous and comparable across different regions. Our findings reveal that the primary factors influencing AA are mainly consistent between urban and rural areas, suggesting that the achievement gap between these environments does not primarily arise from differing educational needs. However, some exceptions exist, such as social support, which is not a significant AA driver in rural areas. This element suggests that for rural students, factors beyond monetary support—such as cultural capital and the perceived value of education—play a more crucial role and are often lacking in rural Portugal and other countries with similar needs. Contrary to most research and expectations in the digital age, our analysis also found that access to ICT does not significantly influence AA, likely because access to devices and the Internet has become nearly universal in Portugal. Instead, how ICT is used—along with factors like ICT literacy, competence, and interest—is likely a more critical determinant of AA, particularly for students from low-income backgrounds who may lack the necessary digital skills and support systems.

Consistent with expectations, academic background variables emerged as the most significant AA drivers. Additionally, the negative impact of grade retention is underscored, emphasizing the need for cost-effective policies to prevent the long-term negative consequences associated with retention. Our results underscore the need for targeted policies addressing rural education's unique challenges, such as improving cultural capital and redefining school catchment areas to promote social mixing. Additionally, our results suggest that policymakers should shift focus from merely providing access to ICT to enhancing the quality of ICT use and digital literacy. The success of programs like Portugal's TEIP shows that targeted interventions can significantly reduce educational disparities and the likelihood of early dropout, but further efforts are needed to ensure equitable funding, promote diverse school environments, and apply lessons from other programs and progressive ruralism to support sustainable development in rural education.

## 7- Declarations

### 7-1- Author Contributions

Conceptualization, A.B.A. and F.C.J.; methodology, A.B.A.; validation, F.C.J.; writing—original draft preparation, A.B.A.; writing—review and editing, F.C.J. All authors have read and agreed to the published version of the manuscript.

### 7-2- Data Availability Statement

The data presented in this study are available on request from the corresponding author.

### 7-3- Funding and Acknowledgement

### 7-4- Institutional Review Board Statement

Not applicable.

### 7-5- Informed Consent Statement

Not applicable.

### 7-6- Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this manuscript. In addition, the ethical issues, including plagiarism, informed consent, misconduct, data fabrication and/or falsification, double publication and/or submission, and redundancies have been completely observed by the authors.

## 8- References

[1] Soares da Silva, D., Figueiredo, E., Eusébio, C., & Carneiro, M. J. (2016). The countryside is worth a thousand words - Portuguese representations on rural areas. Journal of Rural Studies, 44, 77–88. doi:10.1016/j.jrurstud.2016.01.012.

[2] Bachore, M. M. (2022). English Language Literacy Skills and Academic Achievement of Urban and Rural Secondary Schools: The Case of High and Low Achievers. Education Research International, 1, 2315426. doi:10.1155/2022/2315426.

[3] Young, D. J. (1998). Rural and Urban Differences in Student Achievement in Science and Mathematics: A Multilevel Analysis. School Effectiveness and School Improvement, 9(4), 386–418. doi:10.1080/0924345980090403.

[4] Woessmann, L. (2010). Institutional determinants of school efficiency and equity: German states as a microcosm for OECD countries. Jahrbucher Fur Nationalokonomie Und Statistik, 230(2), 234–270. doi:10.1515/jbnst-2010-0206.

[5] Gordon, I., & Monastiriotis, V. (2007). Education, location, education: A spatial analysis of English secondary school public examination results. Urban Studies, 44(7), 1203–1228. doi:10.1080/00420980701302387.

[6] Bratti, M., Checchi, D., & Filippin, A. (2021). Territorial Differences in Italian Students' Mathematical Competencies: Evidence from PISA 2003. SSRN Electronic Journal, IZA Discussion Paper No. 2603. doi:10.2139/ssrn.969356.

[7] Hillier, C., Zarifa, D., & Hango, D. (2022). Mind the Gaps: Examining Youth's Reading, Math and Science Skills across Northern and Rural Canada. Rural Sociology, 87(1), 264–302. doi:10.1111/ruso.12401.

[8] Murphy, S. (2019). School location and socioeconomic status and patterns of participation and achievement in senior secondary mathematics. Mathematics Education Research Journal, 31(3), 219–235. doi:10.1007/s13394-018-0251-9.

[9] Hung, M., Smith, W. A., Voss, M. W., Franklin, J. D., Gu, Y., & Bounsanga, J. (2020). Exploring Student Achievement Gaps in School Districts across the United States. Education and Urban Society, 52(2), 175–193. doi:10.1177/0013124519833442.

[10] Zhao, G., Ye, J., Li, Z., & Xue, S. (2017). How and why do Chinese urban students outperform their rural counterparts? China Economic Review, 45, 103–123. doi:10.1016/j.chieco.2017.06.006.

[11] República Portuguesa (2024). Programa Nacional para a Coesão Territorial - O Interior em números: Território, 35–51. Available online: https://www.portugal.gov.pt/pt/gc21/governo/programa/programa-nacional-para-a-coesao-territorial-/ficheiros-coesao-territorial/programa-nacional-para-a-coesao-territorial-o-interior-em-numeros-os-territorios-do-interior-pdf.aspx (accessed on November 2024).

[12] Marôco, J. (2021). Portugal: The PISA Effects on Education. In: Crato, N. (Eds) Improving a Country's Education. Springer, Cham, Switzerland. doi:10.1007/978-3-030-59031-4_8.

[13] Pordata. (2011). Resident population aged 15 and over according to the Census: total and by highest level of education completed -Municipalities. Pordata, Fundação Francisco Manuel dos Santos, Lisbon, Portugal. Available online: http://www.pordata.pt/Municipios/População+residente+com+15+e+mais+anos+segundo+os+Censos+total+e+por+nível+de+escolaridade+completo+mais+elevado-69 (accessed on November 2024).

[14] Lee, H. (2022). What drives the performance of Chinese urban and rural secondary schools: A machine learning approach using PISA 2018. Cities, 123, 103609. doi:10.1016/j.cities.2022.103609.

[15] Schwerter, J., Bleher, J., Doebler, P., & McElvany, N. (2023). Metropolitan, Urban, and Rural Regions – How Regional Differences Affect Elementary School Students' Academic Achievement, Well-Being, and Motivation. SSRN Electronic Journal. doi:10.2139/ssrn.4368170.

[16] Marker, C., Gnambs, T., & Appel, M. (2018). Active on Facebook and Failing at School? Meta-Analytic Findings on the Relationship between Online Social Networking Activities and Academic Achievement. Educational Psychology Review, 30(3), 651–677. doi:10.1007/s10648-017-9430-6.

[17] Pereira, M. C., & Reis, H. (2012). Diferenças regionais no desempenho dos alunos portugueses: Evidencia do programa PISA da OCDE. Boletim Económico de Inverno, Banco de Portugal, 59–83.

[18] Genesee, F., Lindholm-Leary, K., Saunders, B., & Christian, D. (2006). Educating English Language Learners. In Educating English Language Learners. Cambridge University Press, Cambridge, United Kingdom. doi:10.1017/cbo9780511499913.

[19] Salmela-Aro, K., & Tynkkynen, L. (2012). Gendered pathways in school burnout among adolescents. Journal of Adolescence, 35(4), 929–939. doi:10.1016/j.adolescence.2012.01.001.

[20] Cruz-Jesus, F., Castelli, M., Oliveira, T., Mendes, R., Nunes, C., Sa-Velho, M., & Rosa-Louro, A. (2020). Using artificial intelligence methods to assess academic achievement in public high schools of a European Union country. Heliyon, 6(6), 4081. doi:10.1016/j.heliyon.2020.e04081.

[21] Asif, R., Merceron, A., Ali, S. A., & Haider, N. G. (2017). Analyzing undergraduate students' performance using educational data mining. Computers and Education, 113, 177–194. doi:10.1016/j.compedu.2017.05.007.

[22] Cohn, E., Cohn, S., Balch, D. C., & Bradley, J. (2004). Determinants of undergraduate GPAs: SAT scores, high-school GPA and high-school rank. Economics of Education Review, 23(6), 577–586. doi:10.1016/j.econedurev.2004.01.001.

[23] Miguéis, V. L., Freitas, A., Garcia, P. J. V., & Silva, A. (2018). Early segmentation of students according to their academic performance: A predictive modelling approach. Decision Support Systems, 115, 36–51. doi:10.1016/j.dss.2018.09.001.

[24] Nunes, C., Oliveira, T., Castelli, M., & Cruz-Jesus, F. (2023). Determinants of academic achievement: How parents and teachers influence high school students' performance. Heliyon, 9(2). doi:10.1016/j.heliyon.2023.e13335.

[25] Holmes, C. T. (1989). Grade level retention effects: A meta-analysis of research studies. Flunking grades: Research and Policies on Retention, 16, 33.

[26] Klapproth, F., Schaltz, P., Brunner, M., Keller, U., Fischbach, A., Ugen, S., & Martin, R. (2016). Short-term and medium-term effects of grade retention in secondary school on academic achievement and psychosocial outcome variables. Learning and Individual Differences, 50, 182–194. doi:10.1016/j.lindif.2016.08.014.

[27] Pagani, L., Tremblay, R. E., Vitaro, F., Boulerice, B., & McDuff, P. (2001). Effects of grade retention on academic performance and behavioral development. Development and Psychopathology, 13(2), 297–315. doi:10.1017/S0954579401002061.

[28] Lounkaew, K. (2013). Explaining urban-rural differences in educational achievement in Thailand: Evidence from PISA literacy data. Economics of Education Review, 37, 213–225. doi:10.1016/j.econedurev.2013.09.003.

[29] Dale, P. S., Paul, A., Rosholm, M., & Bleses, D. (2023). Prediction from early childhood vocabulary to academic achievement at the end of compulsory schooling in Denmark. International Journal of Behavioral Development, 47(2), 123–134. doi:10.1177/01650254221116878.

[30] Archibald, S. (2006). Narrowing in on educational resources that do affect student achievement. Peabody Journal of Education, 81(4), 23–42. doi:10.1207/s15327930pje8104_2.

[31] Hodis, F. A., Johnston, M., Meyer, L. H., McClure, J., Hodis, G. M., & Starkey, L. (2015). Maximal levels of aspiration, minimal boundary goals, and their relationships with academic achievement: The case of secondary-school students. British Educational Research Journal, 41(6), 1125–1141. doi:10.1002/berj.3189.

[32] Costa-Mendes, R., Oliveira, T., Castelli, M., & Cruz-Jesus, F. (2021). A machine learning approximation of the 2015 Portuguese high school student grades: A hybrid approach. Education and Information Technologies, 26(2), 1527–1547. doi:10.1007/s10639-020-10316-y.

[33] Valli Jayanthi, S., Balakrishnan, S., Lim Siok Ching, A., Aaqilah Abdul Latiff, N., & Nasirudeen, A. M. A. (2014). Factors Contributing to Academic Performance of Students in a Tertiary Institution in Singapore. American Journal of Educational Research, 2(9), 752–758. doi:10.12691/education-2-9-8.

[34] Musso, M. F., Hernández, C. F. R., & Cascallar, E. C. (2020). Predicting key educational outcomes in academic trajectories: a machine-learning approach. Higher Education, 80(5), 875–894. doi:10.1007/s10734-020-00520-7.

[35] Long, K., & Renbarger, R. (2023). Persistence of Poverty: How Measures of Socioeconomic Status Have Changed Over Time. Educational Researcher, 52(3), 144–154. doi:10.3102/0013189X221141409.

[36] Steinmayr, R., Dinger, F. C., & Spinath, B. (2010). Parents' education and children's achievement: The role of personality. European Journal of Personality, 24(6), 535-550. doi:10.1002/per.755.

[37] Coleman, J. S. (1968). Equality of Educational Opportunity. Equity and Excellence in Education, 6(5), 19–28. doi:10.1080/0020486680060504.

[38] Tesfagiorgis, M., Tsegai, S., Mengesha, T., Craft, J., & Tessema, M. (2020). The correlation between parental socioeconomic status (SES) and children's academic achievement: The case of Eritrea. Children and Youth Services Review, 116(July), 105242. doi:10.1016/j.childyouth.2020.105242.

[39] Moulin, L. (2023). Do private schools increase academic achievement? Evidence from France. Education Economics, 31(2), 247–274. doi:10.1080/09645292.2022.2061428.

[40] Acuña, E., & Rodriguez, C. (2004). The Treatment of Missing Values and its Effect on Classifier Accuracy. Classification, Clustering, and Data Mining Applications, 639–647. doi:10.1007/978-3-642-17103-1_60.

[41] Bilal, M., Omar, M., Anwar, W., Bokhari, R. H., & Choi, G. S. (2022). The role of demographic and academic features in a student performance prediction. Scientific Reports, 12(1), 12508. doi:10.1038/s41598-022-15880-6.

[42] PDR (2020). Lista de Freguesias Rurais PDR 2020. Programa de Desenvolvimento Rural 2014-2020, Lisbon, Portugal.

[43] Ricciardi, C., Manfra, L., Hartman, S., Bleiker, C., Dineheart, L., & Winsler, A. (2021). School readiness skills at age four predict academic achievement through 5th grade. Early Childhood Research Quarterly, 57, 110–120. doi:10.1016/j.ecresq.2021.05.006.

[44] Wakefield, J., & Frawley, J. K. (2020). How does students' general academic achievement moderate the implications of social networking on specific levels of learning performance? Computers and Education, 144. doi:10.1016/j.compedu.2019.103694.

[45] James, G., Witten, D., Hastie, T., & Tibshirani, R. (2000). An introduction to Statistical Learning. Current Medicinal Chemistry, 7(10), 995–1039. doi:10.1007/978-1-4614-7138-7.

[46] Aaronson, D., Barrow, L., & Sander, W. (2007). Teachers and student achievement in the Chicago public high schools. Journal of Labor Economics, 25(1), 95–135. doi:10.1086/508733.

[47] Berthelon, M., Bettinger, E., Kruger, D. I., & Montecinos-Pearce, A. (2019). The Structure of Peers: The Impact of Peer Networks on Academic Achievement. Research in Higher Education, 60(7), 931–959. doi:10.1007/s11162-018-09543-7.

[48] Bowers, A. J. (2010). Grades and graduation: A longitudinal risk perspective to identify student dropouts. Journal of Educational Research, 103(3), 191–207. doi:10.1080/00220670903382970.

[49] Wang, N. (2024). The Urban–Rural Heterogeneous Effect of Family SES on Achievement: The Mediating Role of Culture. Behavioral Sciences, 14(2), 84. doi:10.3390/bs14020084.

[50] Ferrão, M., & Almeida, L. (2019). Differential effect of university entrance score on first-year students' academic performance in Portugal. Assessment and Evaluation in Higher Education, 44(4), 610–622. doi:10.1080/02602938.2018.1525602.

[51] Barbosa-Camacho, F. J., Romero-Limón, O. M., Ibarrola-Peña, J. C., Almanza-Mena, Y. L., … González-Ojeda, A. (2022). Depression, anxiety, and academic performance in COVID-19: a cross-sectional study. BMC Psychiatry, 22(1), 1–10. doi:10.1186/s12888-022-04062-3.

[52] Guèvremont, A., Roos, N. P., & Brownell, M. (2007). Predictors and Consequences of Grade Retention: Examining Data from Manitoba, Canada. Canadian Journal of School Psychology, 22(1), 50–67. doi:10.1177/0829573507301038.

[53] Nunes, C., Beatriz-Afonso, A., Cruz-Jesus, F., Oliveira, T., & Castelli, M. (2022). Mathematics and Mother Tongue Academic Achievement: A Machine Learning Approach. Emerging Science Journal, 6(special issue), 137–149. doi:10.28991/ESJ-2022-SIED-010.

[54] Kubey, R. W., Lavin, M. J., & Barrows, J. R. (2001). Internet use and collegiate academic performance decrements: Early findings. Journal of Communication, 51(2), 366–382. doi:10.1111/j.1460-2466.2001.tb02885.x.

[55] Mensah, F. K., & Kiernan, K. E. (2010). Gender differences in educational attainment: Influences of the family environment. British Educational Research Journal, 36(2), 239–260. doi:10.1080/01411920902802198.

[56] Delen, D. (2010). A comparative analysis of machine learning techniques for student retention management. Decision Support Systems, 49(4), 498–506. doi:10.1016/j.dss.2010.06.003.

[57] Chesters, J., & Daly, A. (2017). Do peer effects mediate the association between family socio-economic status and educational achievement? Australian Journal of Social Issues, 52(1), 65–77. doi:10.1002/ajs4.3.

[58] Bartels, R., & Goodhew, J. (1981). The Robustness of the Durbin-Watson Test. In The Review of Economics and Statistics (Vol. 63, Issue 1, p. 136). doi:10.2307/1924228.

[59] Anscombe, F. J., & Tukey, J. W. (1963). The Examination and Analysis of Residuals. Technometrics, 5(2), 141–160. doi:10.1080/00401706.1963.10490071.

[60] O'Brien, R. M. (2007). A caution regarding rules of thumb for variance inflation factors. Quality and Quantity, 41(5), 673–690. doi:10.1007/s11135-006-9018-6.

[61] Belsley, D. a, Kuh, E., & Welsch, R. E. (1980). Identifying influential data and sources of collinearity. Wiley, New Jersey, United States. doi:10.1002/0471725153.

[62] White, H. (1980). A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity. Econometrica, 48(4), 817. doi:10.2307/1912934.

[63] Dean, J., Roberts, P., Downes, N., & Goldsmith, A. (2023). The spatial implications of academic achievement in Year 12: Rethinking discourses of disadvantage in rural locations. Australian Journal of Education, 67(1), 94–114. doi:10.1177/00049441231155708.

[64] Mariano, L. T., Martorell, P., & Berglund, T. (2024). The effects of grade retention on high school outcomes: Evidence from New York City schools. Journal of Research on Educational Effectiveness, 1-31. doi:10.1080/19345747.2023.2287607.

[65] Nunes, L. C., Balcão Reis, A., & Seabra, C. (2018). Is retention beneficial to low-achieving students? Evidence from Portugal. Applied Economics, 50(40), 4306–4317). doi:10.1080/00036846.2018.1444261.

[66] Larsen, M. F., & Valant, J. (2024). The long-term effects of grade retention: Evidence on persistence through high school and college. Journal of Research on Educational Effectiveness, 17(4), 615-646. doi:10.1080/19345747.2023.2240323.

[67] Martin, A. J. (2011). Holding back and holding behind: Grade retention and students' non-academic and academic outcomes. British Educational Research Journal, 37(5), 739–763. doi:10.1080/01411926.2010.490874.

[68] Hill, A. J. (2014). The costs of failure: Negative externalities in high school course repetition. Economics of Education Review, 43, 91–105. doi:10.1016/j.econedurev.2014.10.002.

[69] Freeman, J., & Simonsen, B. (2015). Examining the Impact of Policy and Practice Interventions on High School Dropout and School Completion Rates: A Systematic Review of the Literature. Review of Educational Research, 85(2), 205–248. doi:10.3102/0034654314554431.

[70] Määttä, M., & Aaltonen, S. (2016). Between rights and obligations – rethinking youth participation at the margins. International Journal of Sociology and Social Policy, 36(3–4), 157–172. doi:10.1108/IJSSP-09-2014-0066.

[71] Mazrekaj, D., & De Witte, K. (2020). The effect of modular education on school dropout. British Educational Research Journal, 46(1), 92–121. doi:10.1002/berj.3569.

[72] DataReportal. (2022). Digital 2022: Portugal. DataReportal – Global Digital Insights: Kepios. Available online: https://datareportal.com/reports/digital-2022-portugal (accessed on November 2024).

[73] Eurostat (2023). ICT usage in households and by individuals (isoc_i). Eurostat Metadata, Luxembourg City, Luxembourg Available online: https://ec.europa.eu/eurostat/cache/metadata/EN/isoc_i_simsih2_pt.htm (accessed on November 2024).

[74] Ministério do Planeamento (2021). Recover Portugal, Rebuilding the Future: Recovery and Resilience Plan. Available online: https://www.consultalex.gov.pt/ConsultaPublica_Detail.aspx?Consulta_Id=183 (accessed on November 2024).

[75] Lei, H., Xiong, Y., Chiu, M. M., Zhang, J., & Cai, Z. (2021). The relationship between ICT literacy and academic achievement among students: A meta-analysis. Children and Youth Services Review, 127. doi:10.1016/j.childyouth.2021.106123.

[76] Park, S., & Weng, W. (2020). The relationship between ICT-related factors and student academic achievement and the moderating effect of country economic indexes across 39 countries: Using multilevel structural equation modelling. Educational Technology and Society, 23(3), 1–15. doi:10.2307/26926422.

[77] Ren, W., Zhu, X., & Yang, J. (2022). The SES-based difference of adolescents' digital skills and usages: An explanation from family cultural capital. Computers and Education, 177. doi:10.1016/j.compedu.2021.104382.

[78] Verbree, A. R., Hornstra, L., Maas, L., & Wijngaards-de Meij, L. (2023). Conscientiousness as a Predictor of the Gender Gap in Academic Achievement. Research in Higher Education, 64(3), 451–472. doi:10.1007/s11162-022-09716-5.

[79] King, R. B. (2016). Gender differences in motivation, engagement and achievement are related to students' perceptions of peer—but not of parent or teacher—attitudes toward school. Learning and Individual Differences, 52, 60–71. doi:10.1016/j.lindif.2016.10.006.

[80] Wang, J., & Wu, Y. (2023). Income inequality, cultural capital, and high school students' academic achievement in OECD countries: A moderated mediation analysis. British Journal of Sociology, 74(2), 148–172. doi:10.1111/1468-4446.12997.

[81] De Graaf, N. D., De Graaf, P. M., & Kraaykamp, G. (2000). Parental cultural capital and educational attainment in the Netherlands: A refinement of the cultural capital perspective. Sociology of Education, 73(2), 92–111. doi:10.2307/2673239.

[82] Wang, W., Dong, Y., Liu, X., Bai, Y., & Zhang, L. (2020). The effect of parents' education on the academic and non-cognitive outcomes of their children: Evidence from China. Children and Youth Services Review, 117. doi:10.1016/j.childyouth.2020.105307.

[83] Trinidad, J. E. (2020). Will it matter who I'm in school with? Differential influence of collective expectations on urban and rural US schools. International Studies in Sociology of Education, 29(4), 344–365. doi:10.1080/09620214.2019.1673791.

[84] Agger, C., Meece, J., & Byun, S. Yong. (2018). The Influences of Family and Place on Rural Adolescents' Educational Aspirations and Post-secondary Enrollment. Journal of Youth and Adolescence, 47(12), 2554–2568. doi:10.1007/s10964-018-0893-7.

[85] Banerjee, P. A. (2016). A systematic review of factors linked to poor academic performance of disadvantaged students in science and maths in schools. Cogent Education, 3(1), 1178441. doi:10.1080/2331186X.2016.1178441.

[86] Chesters, J., & Cuervo, H. (2022). (In)equality of opportunity: educational attainments of young people from rural, regional and urban Australia. Australian Educational Researcher, 49(1), 43–61. doi:10.1007/s13384-021-00432-0.

[87] Oberti, M., & Savina, Y. (2019). Urban and school segregation in Paris: The complexity of contextual effects on school achievement: The case of middle schools in the Paris metropolitan area. Urban Studies, 56(15), 3117–3142. doi:10.1177/0042098018811733.

[88] Golsteyn, B. H. H., Non, A., & Zölitz, U. (2021). The impact of peer personality on academic achievement. Journal of Political Economy, 129(4), 1052–1099. doi:10.1086/712638.

[89] Martínez Abad, F., & Chaparro Caso López, A. A. (2017). Data-mining techniques in detecting factors linked to academic achievement. School Effectiveness and School Improvement, 28(1), 39–55. doi:10.1080/09243453.2016.1235591.

[90] Serrati, P. S. (2024). School and residential segregation in the reproduction of urban segregation: A case study in Buenos Aires. Urban Studies, 61(2), 313–330. doi:10.1177/00420980231178401.

[91] Damme, M.-C. O. J. Van. (2001). Relationship between School Composition and Characteristics of School Process and their Effect on Mathematics Achievement. British Educational Research Journal, 27(4), 407–432. doi:10.1080/01411920125573.

[92] Kallert, A., Belina, B., Miessner, M., & Naumann, M. (2021). The Cultural Political Economy of rural governance: Regional development in Hesse (Germany). Journal of Rural Studies, 87, 327–337. doi:10.1016/j.jrurstud.2021.09.017.

[93] Abrantes, P., Mauritti, R., Roldão, C., Alves, L., Amaral, P., Baptista, I., ... & Teixeira, A. (2011). Effects of TEIP: An Evaluation of the Academic and Social Impacts in Seven Territories of Educational Priority Intervention. A Synthesis of Results, Lisbon, Portugal.

[94] Pezzoli, K., Hibbard, M., & Huntoon, L. (2009). Introduction to symposium: Is progressive regionalism an actionable framework for critical planning theory and practice? Journal of Planning Education and Research, 28(3), 336–340. doi:10.1177/0739456X08329711.