Volume measures for non-market education output[1]

Jeremy Fouliard (INSEE), Mats Le Floch (INSEE), Pauline Mendras (INSEE)

Abstract

Measuring the volume of public services output is a fundamental challenge for national accounts, due to the absence of market prices. In France, the volume of non-market education is measured using an output approach based on student enrollment. As this indicator mainly reflects demographic trends, the recent decline in the number of students mechanically reduces the estimated volume of education, even though quality could have improved through smaller class sizes. Conversely, the decline in PISA scores could reflect a decrease in teaching quality. This study compares alternative volume measurements, and examines different quality adjustments to capture other features of the service provided. It introduces a theoretical framework that incorporates quality adjustments into output volume measures based on the contribution of quality to human capital, drawing on estimates from the literature. Between 2013 and 2022, the adjustment based on PISA scores lowers annual growth rates (-0.056 percentage points on average with a maximum of -0.1 points) while the class-size adjustment slightly raises annual growth (+0.015 percentage points on average, peaking at +0.035 points). Adjusting for the share of novice teachers has a negligible effect (below 0.01 points). Overall, quality adjustments have a modest effect on the estimated volume of education output, reflecting the low sensitivity of future earnings to the quality indicators considered. However, those moderate effects could accumulate over time and result in visible change in trend. Indeed, a stratification of the education service to account for the rising enrollment of pupils with disability increases output volume growth rates by 0.15 points at most, but leads to a cumulative change of about one percentage point over the period 2006–2023.

[1]We are very grateful to Philippe Askenazy, Cédric Audenis, Alexandre Bourgeois, Abe Dunn, Maryse Fesseau, Pauline Givord, Jean-Cyprien Heam, Guillaume Houriez, Marie Leclair, Matthieu Lequien, Sylvie Rousseau, Corentin Trevien and Meryam Zaiem for their useful feedback. We are highly indebted to Richard Heys and Paul Schreyer for their generous help. Our thanks also go to the seminar participants at the ESCoE Conference 2026, GENA meeting 2026, IARIW Conference 2026, the INSEE Research Seminar, Journées de la Méthodologie Statistique 2025,UNNES Sprint 2026 and all the members of the National Accounts Department at INSEE and the Statistical Office of the French Ministry of Education (DEPP) for their help with the data.

1 Introduction

Measuring the production of public goods and services remains a major challenge in economics. As highlighted in the pioneering Atkinson report (Atkinson, 2005), imprecise measurement of public service output represents a critical issue for economic analysis. It can distort evaluations of the output gap and resource utilization, ultimately influencing fiscal and monetary policy decisions. Furthermore, without accurate measurement, national accounts fail to fully capture the contribution of public services to long-term economic growth, which is mainly driven by human capital accumulation according to the seminal work of Lucas Jr (1988). In response to this challenge, a key role of national accounts is to decompose nominal aggregates into two components: a price index, capturing price dynamics, and a volume index, measuring changes in real quantities. This so-called volume–price split is a fundamental tool in economic analysis. It enables the measurement of real output growth as well as detailed evaluations of factor productivity and its dynamics.

However, a key issue is that most public goods and services are non-market and therefore lack prices that would allow for direct measurement of output volume. This issue is particularly salient for the education sector. Even if it accounts for only 3% of total production in 2023, it represents one of the largest components of non-market output, accounting for roughly one-third of non-market production in 2023 (Insee, 2025), with educational services provided either free of charge or heavily subsidized. Because educational services lack prices, their value is defined as the sum of costs incurred in their production[2] following the System of National Accounts (SNA 2025). Three main types of methods are generally employed to estimate their production volume. First, the input-based method consists in applying the volume–price split to the production factors rather than to the final output. Elementary price indices are estimated for each production factor (labor, consumption of capital and intermediate consumption). These indices are then aggregated into an overall price index by weighting each factor according to its share in total production costs. The production volume is then obtained by dividing the production value by this price index. As recommended by international guidelines, this method is notably used by Insee to measure the production of collective public services (defense, justice, etc.). However, this method is not recommended for education under EU statistical conventions as it shows no significant change in public service productivity: by definition, the output volume is equal to the quantity of inputs used.

Second, the output-based method consists in using a non-monetary variable that measures the production of the service as an indicator of volume. Several measures can be used to capture the quantity of the service, such as the number of teaching hours attended by students in education or the number of treatments provided to patients in the case of health services. This method is currently used in French national accounts to measure teaching activity based on the number of student attendance days[3], broken down into nine educational levels. These numbers are weighted by the public expenditure allocated to each level of education, in order to account for disparities in the cost per student. This approach captures the quantity of services delivered to users, but abstracts from quality improvements that are not reflected in the stratification. As a result, a drop in the volume indicator may not signal a decline in service production if services have become more efficient or better tailored to users' needs.

The third methodological approach, known as the outcome-based method, emphasizes the impact of the service on a predefined target variable. Unlike input- or output-focused approaches, this method prioritizes measurable results, assessing effectiveness based on achieved outcomes rather than resource allocation or service volume. In the case of education, this approach may view teaching services as

[2]The production of public education services is measured by the National Accounts Department (DCN) of the French Statistical Office (Insee). It is complemented by a thematic account for education developed by the French Ministry of Education, which provides more detailed information on producers and the different levels of education.

[3]During the Covid period, an adjustment linked to the volume of teaching hours was performed. The number of student attendance days is fixed each year according to the number of holidays. It could change in case of a structural reform

an investment in human capital, whose value can be estimated from the student's future earnings. More broadly, education services could also be assessed in terms of other dimensions, such as students' well-being, social integration or the development of civic skills. However, the European System of Accounts (ESA 2010) forbids this method, due to the difficulty of identifying the exact contribution of the education system to an individual's career outcomes.

The output-based volume measure currently estimated by Insee is primarily driven by demographic changes, although it can also reflect other factors such as compositional shifts across fields of study in higher education. As a result, productivity measurements are mainly determined by the disparity between the dynamics of the student population and those of educational staff. This approach, however, offers no insight into the quality of teaching provided. Thus, under the assumption of constant costs, demographic projections indicating a decline in student numbers would mechanically result in a reduction in the measured volume of education. This decrease would occur even if the quality of teaching services improves, for instance as a result of smaller class sizes, or it the quality worsens as suggested by the recent PISA tests results for France. For this reason, several studies suggest alternative indicators aimed at better capturing variations in service quality (Schreyer, 2010; Gemmell et al., 2017; ONS, 2025). In addition, the System of National Accounts (2025)[4] endorses their use, although the European System of Accounts (2010)[5] prohibits quality-adjusted estimation methods to ensure cross-country comparability.

This study introduces a novel approach to computing volume and price indices for non-market education services, focusing on first and second degree education[6], within a theoretical framework that explicitly incorporates quality adjustments, emphasizing the contribution of quality to human capital, and builds on estimates that are consistent with the existing literature. The issue of quality is especially salient in the non-market sector. While, in the market sector, volumes are typically obtained by deflating nominal output with constant-quality price indices, such an approach is not available when volume is measured directly, making quality assessment inherently more complex. By treating education as an investment in human capital, this approach incorporates an outcome component into the output volume measure, highlighting how improvements in school quality translates into productivity gains. Using French national accounts data supplemented by series from the Ministry of Education, it analyzes both input and output-based methods. It also compares various adjustments designed to better capture quality changes, including explicit quality adjustments, alternative quantity measures, and a finer stratification, in order to contribute to international discussions on improving methodologies for volume and price measurement[7].

The quality of education services is both an abstract and multidimensional concept. It can be defined with reference to the objectives of the education system, which may include several aspects: delivering fundamental knowledge, promoting equal opportunities and inclusion, ensuring labor market integration, and fostering citizenship. In the French context, these purposes are explicitly defined and codified within the provisions of the Code of Education[8]. These goals can be approached through several metrics, related to the inputs of the production process (teacher qualifications, and experience),

[4] SNA 2025, Paragraph 18.130 "In the light of these observations, the "output volume method" is the recommended method for compiling indicators of volume change of non-market services. The method is based on quantity indicators, adequately quality-adjusted, weighted together using average cost weights."

[5] ESA 2010, Paragraph 10.30 : "In the European Union, given the conceptual difficulties and the absence of consensus on output methods adjusted for quality (based on outcome), such methods are excluded from the central framework in order to preserve the comparability of the results. Such methods are reserved on an optional basis for supplementary tables, while continuing research. Thus, in the field of non-market health and education, the estimates of production and of consumption in volume terms have to be calculated on the basis of direct output measures - not adjusted for quality - by weighing up the quantities produced by the previous year unit costs of those services, without applying any correction to them in order to take account of quality"

[6] i.e., excluding higher education

[7] Table B in the appendix summarizes the tested methods, estimated volumes, data sources, estimation period, and stratification used.

[8] Articles L121-1 to L124-20 of Title II, "Objectives and missions of the public education service," Code of Education

to the service itself (class size, school climate), or to the outcomes of this production (assessment results, exam success rates). These dimensions are used here to build different quality adjustments to the education volume measurement. Quality change can also be captured through a detailed classification of education services that distinguishes student populations based on their different teaching arrangements and support structures, for example pupils with disabilities.

Using a finer stratification of student populations with specific strata for students with disabilities in each level have a moderate effect on growth rates each year (+0.15 percentage points at most in 2023), but the gaps accumulate over time as the enrollment of pupils with disability is constantly growing since 2005. As a result, the overall relative change from 2006 to 2023 goes from -0.5% for the unadjusted volume to 0.5% with the more detailed stratification. This adjustment improves the measurement of education output because it better captures the distinct nature of the educational services provided to these students, linked to higher cost per pupil. Hence, it can be considered as an implicit quality adjustment.

The explicit quality adjustments leveraging quality metrics based on their contribution to human capital have a low impact on the measured volume of education output. Between 2013 and 2022, the differences in growth rates between the quality-adjusted volumes and the unadjusted volume are modest or negligible, with the largest annual difference being 0.1 percentage points for the PISA scores adjustment. During this period, the volume adjusted for PISA scores grew more slowly than the unadjusted volume, while the volume adjusted for class size increased faster. This difference can be explained by the evolution of the two adjustment variables: on the one hand, average class size has decreased since 2017 due to demographic decline and the targeted class size reduction of classes in priority education zones; on the other hand, average PISA scores globally follow a declining trend since 2012. Finally, adjusting the education volume for the proportion of novice teachers results in volume growth estimates that are very similar to those of the unadjusted volume. The small differences between quality-adjusted and unadjusted volumes are mostly explained by the limited impact of quality indicators on students' future earnings estimated in the empirical literature, and partly because some adjustments apply only to certain educational levels. As a result, implementing these quality adjustments in national accounts would have a limited effect on the estimated volume of education output, as well as on related aggregates such as value added. These findings could nonetheless change if new estimates of the effects of quality indicators are available in the future, or if quality indicators undergo substantial changes. In addition, small change in growth rates could lead to substantial cumulative change over time if the quality indicators follow a sustained upward or downward trend in the future.

The modest effect of quality on variations in volume means that it explains only a small share of the change in human capital compared to the evolution of headcounts: broadly, over the last decade, quality changes affect future earnings and thereby volume by at most 0.05 percentage points per year (in absolute value), while the rate of change in the number of pupils is around 0.1 to 0.5 percentage points in most years (in absolute value), so it is ten times larger. Nonetheless, this does not imply that quality has no effect, or that policies aimed at improving it are useless. Indeed, Fajeau et al. (2025b) shows, using similar empirical estimates, that the welfare benefits of reducing class sizes may exceed the net cost including future taxes on pupils' earnings, provided the calculation is based on a sufficiently long time horizon. This is because a small *relative* increase in the total discounted sum of wages of a pupil is still a large amount of money compared to a teacher's wage, and because the high marginal tax rate induces future gains for the government. This holds despite the fact that a class size reduction decreases productivity, since the *relative* change in the number of teachers is higher than that of future wages.

Section 2 reviews the literature on quality adjusting measurements of non-market services. Section 3 develops a theoretical framework for the different volume measures and quality adjustments. Section 4 applies the input-based method, as well as several variants of this approach. Section 5 presents

different output-based measures of education volume, expressed in terms of number of pupils and number of pupil-hours. Section 6 introduces a stratification suited for students with disabilities. Section 7 examines quality-adjusted volume measures based on indicators related to the service itself (class size, school climate), the inputs (share of novice teachers) and the results of the education service (standardized test scores). Section 8 summarizes and compares the results. Section 9 concludes with suggestions for future research.

2 Literature review

A substantial body of work conducted by academic studies and national statistical institutes has examined the measurement of non-market services, concerning education (Ahlroth et al. (1997) , Gemmell et al. (2017)) or health (Cutler et al. (2022), Weale (2025)). Among the most influential contributions Diewert (2008) and Schreyer (2012) introduce theoretical frameworks that link volume and price measurement to standard microeconomic models and discuss possible productivity measurements. Schreyer (2012) presents different ways to apply implicit and explicit quality adjustments. He suggests integrating quality into volume measurement according to its contribution to an outcome variable - human capital for education, life years for health. He also proposes using regressions of costs on quality variables, an approach analogous to hedonic regressions commonly used for market products. His study emphasizes the importance of using an appropriate stratification which reflects consumers' utility. In a previous work, Schreyer (2010) identifies three principal approaches to measure quality variations: i) a finer stratification, ii) an explicit quality adjustment and iii) an outcome-based volume measure base for example on students' future earnings. Using an explicit quality adjustment is useful to take into account dimensions of service quality that are not captured by stratification, with potential indicators including exam results, class size, and inspection reports.

Several indicators have been explored to adjust education volume, with student performance in national assessments being the most widely used. The UK Office for National Statistics (ONS) applies a +0.25% adjustment to pupil-weighted costs for primary and secondary education based on four-year trends in scores at the General Certificate of Secondary Education examination (GSCE). The ONS also draws on GCSE results to adjust the volume of education for quality and construct an annual education productivity indicator (ONS, 2025). These quality-adjusted education volume and productivity estimates are published separately from the national accounts, as part of a dedicated quarterly publication on public service productivity. PISA scores or graduation rates can also be used, though they can produce divergent results (Gemmell et al., 2017). While most quality adjustments incorporating exam results are applied to an output volume measure, some studies like Gu and Wong (2012) adjust an income-based output measure, by weighting test score improvements according to their estimated impact on future earnings.

Class size is another frequently used indicator. Studies examining the US (Fraumeni et al., 2008) and Italy (Collesi et al., 2007) employ student-teacher ratios as proxies to adjust estimated education volume. Additionally, integrated approaches—combining assessment-based and class size metrics—have been proposed by Hanushek and Ettema (2017). The authors suggest measuring educational productivity using the number of teachers per student (the inverse of class size) as the denominator, and the NAEP[9] Long Term Trend test scores as the numerator.

School inspection reports, as suggested by Schreyer (2010), could be used to construct quality adjustments since they are valuations carried out in schools to assess the quality of teaching, the organization of the school and students' learning outcomes. They could provide useful information by identifying educational outcomes relative to characteristics valued by users, such as teaching

[9]The Long-Term Trend National Assessment of Educational Progress assesses reading and mathematics skills among US students aged 9, 13, and 17

effectiveness or compliance with curriculum standards. However, these indicators are multidimensional, largely subjective, and often qualitative. For this reason, most studies do not incorporate inspection reports.

Other potential quality indicators include teacher composition, repetition rates (Collesi et al., 2008), dropout rates (Fraumeni et al., 2008), and school climate via harassment measures (ONS, 2025). Some studies weight full-time equivalent (FTE) teachers by salary or qualifications to reflect teaching quality (Gemmell et al., 2017). Similarly, Fraumeni et al. (2008) introduce a teacher composition index that integrates both education level and experience, factors which are likely to influence teaching quality. For higher education, teaching and research activities are distinguished (Gemmell et al., 2017): teaching is evaluated via success rates weighted by qualifications (NZQA credits), and research via publications weighted by citations.

In line with outcome-based methods discussed by Schreyer (2010), a human capital approach is also tested by Gemmell et al. (2017), who calculate expected future earnings following education using the *New Zealand Income Survey*. A key limitation is that observed earnings result not only from education but also from other factors such as individual skills, family background and economic conditions. In practice, Gemmell et al. (2017) find that the outcome-based method yields lower growth rates of education production than the output-based method adjusted for graduation rates. The authors explain this gap by the negative impact of rising unemployment and stagnant wages following the 2008 financial crisis on the earnings indicator, whereas the proportion of students obtaining a diploma increased over the same period. Twenty years earlier, Ahlroth et al. (1997) proposed a new output measure for education by applying the Jorgenson and Fraumeni (1989) framework to Sweden. This approach valued education output directly at the increment to pupils' lifetime incomes. Their findings indicate that the estimate would have a substantial impact on GDP growth, though their estimates depend on the wage structure.

Overall, these studies highlight the importance of examining and comparing results across methods and quality adjustments, as their effects on education volume can vary substantially. For instance, Gu and Wong (2012) show that output-based methods without quality adjustment and outcome-based methods produce similar growth rates but very different education production levels. According to Fraumeni et al. (2008), adjusting for class size or repetition rates leads to higher growth rates of volume than using teacher composition indices for the US between 1990 and 2001. Comparing estimation methods is particularly important during crises, when public service provision is hampered. For example, Dkhissi et al. (2025) show that during COVID-19, the estimation methods used in OECD countries significantly affected measured growth of non-market health services but had little effect on education value added, possibly because educational services continued to be provided, albeit often remotely. Finally, quality adjustments of non market output are key nowadays because they service as input into broader measurement frameworks. Heys and Taylor (2025) construct gross and net inclusive income aggregates for the UK, extending the production and asset boundaries to encompass unpaid household services, natural capital and human capital, and take quality-adjusted non-market output as one of their four building blocks. Establishing how quality should enter the volume measure is therefore a prerequisite for such frameworks

3 Conceptual framework

We draw on the conceptual framework and terminology developed by Schreyer (2012) and Diewert (2008) to analyze the measurement of non-market production and the methodology for quality adjustment.

3.1 Notations and market output setting

First, we define the usual producer side setting, valid for both market and non-market output. There are $1 \leq i \leq N$ products (goods or services) in quantity $y = (y_1, ...y_N)$. Each is produced using a technology S_i defining the set of inputs $x = (x_1, ...x_M)$ that can be used to create quantity y_i of product i. The vector of input price is $w = (w_1, ..., w_M)$ so the cost of a production process is $w \cdot x = \sum_{j=1}^{M} w_j x_j$ where $x \in S_i(y_i)$. The cost function of a product i is the minimum cost of producing y_i with technology S_i defined by $C_i(y_i, w) = \min_{x \in S_i(y_i)} w \cdot x$. Therefore, the average cost is $c_i(y_i, w) = \frac{C_i(y_i, w)}{y_i}$ and the total cost is $C(y, w) = \sum_{i=1}^{N} C_i(y_i, w)$. The optimal combination of inputs reaching the minimum cost $C_i(y_i, w)$ is denoted $X_i(y_i, w) = (X_{i1}, ..., X_{iM})$. The time index of the cost functions C_i^t and c_i^t indicates that they correspond to the minimization problem using technology S_i^t available at time t.

For a market product under perfect competition, producers choose the optimal set of inputs $X_i(y_i, w)$ to produce a quantity y_i, and choose y_i so that marginal cost is equal to market price i.e. $\frac{\partial C_i(y, w_i)}{\partial y_i} = p_i$. Consumers equate their marginal utility to price, so that $\frac{\partial U}{\partial y_i}(y_i) = p_i$. Consequently, market prices used for the standard Laspeyres quantity index Q^t and Paasche price index P^t reflect both supply and demand. With those two indexes, the growth in production value can be broken down into a volume and a price component:

$$\frac{\sum_i y_i^t p_i^t}{\sum_i y_i^{t-1} p_i^{t-1}} = Q^t P^t \quad \text{where} \quad Q^t = \frac{\sum_i y_i^t p_i^{t-1}}{\sum_i y_i^{t-1} p_i^{t-1}}, P^t = \frac{\sum_i y_i^t p_i^t}{\sum_i y_i^t p_i^{t-1}}$$

This decomposition can be applied to inputs as well, provided those inputs are market products:

$$\frac{C^t(y^t, w^t)}{C^{t-1}(y^{t-1}, w^{t-1})} = Q_{input}^t P_{input}^t \quad \text{where} \quad \begin{cases} Q_{input}^t = \dfrac{\sum_i X_i^t(y_i^t, w^t) \cdot w^{t-1}}{\sum_i X_i^{t-1}(y_i^{t-1}, w^{t-1}) \cdot w^{t-1}} \\ P_{input}^t = \dfrac{\sum_i X_i^t(y_i^t, w^t) \cdot w^t}{\sum_i X_i^t(y_i^t, w^t) \cdot w^{t-1}} \end{cases}$$

The input volume index Q_{input}^t captures changes in input quantities and distribution, controlling for the change in input costs. It can serve as the denominator of a productivity measure for both market and non-market activities; for example, if the only input is labor considered as homogeneous, Q_{input}^t is the number of full-time equivalents used in usual productivity measurements.

3.2 Non-market output

Non-market products have zero or insignificant prices. The European System of Accounts (ESA 2010, Paragraph 20.19) defines a "significant price" as one that substantially affects the quantity produced and sold. A usual criterion is that it should cover half the production cost over several years. Therefore, the only information available for a non-market product are unit costs $c_i(y_i, w)$, which are treated as quasi-price (Schreyer, 2012). The international conventions define the value of non-market production as the sum of costs:

$$Value^t = C^t(y^t, w^t) = \sum_i X_i(y_i^t, w^t) \cdot w^t$$

Like Schreyer (2012), we assume for simplicity that the observed cost $w \cdot X_i$ still equals the minimum cost $C_i(y_i, w)$, which implies that the non-market producer operates on the production possibility frontier despite the absence of competition. In other words, the public sector is assumed to produce efficiently its services. This assumption may, however, be contradicted for several reasons (Afonso et al., 2010). Public expenditure can, for instance, be diverted for the benefit of specific pressure groups: education spending may be channeled toward a particular level of educators or a specific group of teachers. Aninat and Bauer (1999) describe the Chilean experience, where an increase in real public

spending on health over several years did not generate any visible or measurable improvement in the quantity or quality of services delivered to users, because the additional funds were absorbed as rents by doctors. Various forms of corruption can also undermine the efficiency of public provision (Shleifer and Vishny, 1993).

From the consumers' point of view, the allocation of public spending partly reflects their preference expressed through voting or other political actions, thus costs also represent the choices of consumers to a certain extent. However, the budget allocation may also reflect the preferences of politicians (Besley and Coate, 1997), or the interest of lobby groups (Grossman and Helpman, 1994; Goldberg and Maggi, 1999). For these reasons, while defining the value of non-market output as the sum of costs is imperfect, it remains the most practical estimate for ensuring the coherence and comparability of national accounts. More informative indicators can nonetheless be used to assess the *volume* of non-market output.

3.2.1 Input volume measures

International conventions (SNA and ESA) recommend using the volume of input Q^t_{input} as the volume of non-market production when better methods cannot be leveraged. This is the case for collective public services like defense, police, general administration and justice. While easy to implement, setting $Q^t = Q^t_{\text{input}}$ implies that productivity is constant by construction. Eurostat (Eurostat, 2016) classifies this method as low-quality (category C) for education and health services. The limits of measuring the value of non-market products as the sum of costs mentioned in the previous section also apply to measuring its volume as the quantity of input: it provides no insight into how inputs are selected, allocated, or combined during the production process, and it may be disconnected from the preferences and welfare of consumers.

Even when the volume of inputs is not used as a measure of output, it remains valuable for other purposes such as building robust productivity estimates. Since productivity is defined as the ratio of output volume to input volume, the way input volume is defined and measured can have a substantial effect on the resulting productivity dynamics. Producing precise input quantity indices is therefore important for economic analysis across all types of activity, including market sectors.

Building an accurate index of input quantities is however not straightforward. Since inputs are typically market products, the challenges involved are those commonly encountered when constructing volume indices more generally, such as the availability of detailed quantity and price data. The usual and most simple way to compute Q^t_{input} is to divide the total cost $C^t(y^t, w^t)$ by an index of input price P^t_{input} using aggregate prices index of labor, intermediate consumption and capital consumption. However, $Q^t_{input} = \frac{\sum_i X_i^t \cdot w^{t-1}}{\sum_i X_i^{t-1} \cdot w^{t-1}}$ can also be obtained directly from input quantities and prices. This approach requires more data, but enables finer stratification and thereby more precise measurement. In particular, weighting the number of full-time equivalents across worker categories by their respective wages accounts for labor composition, and captures some variations of input quality. Gemmell et al. (2017) uses this approach to compare different notions of productivity of non-market education, and the EU-KLEMS database (O'Mahony and Timmer, 2009) relies on this type of input quantity index in its growth accounting framework.

A variant of the input volume measure consists in using the price index of comparable market products P^t_{market} to deflate the total cost and obtain $Q^t_{input-market} = C^t(y^t, w^t)/P^t_{market}$ as in Mitchell et al. (2022). Market prices better reflect supply costs and the preferences of users, but non-market services like primary education rarely have comparable market counterpart.

3.2.2 Output volume measures

Eurostat's handbook (Eurostat, 2016) recommends measuring the volume of public services by relying on a *quantity variable* when it is possible. The quantity variable should reflect the intertemporal utility provided by the service to the user. For education services, Eurostat indicates that the number of pupil-hours[10] should be used as the quantity. The amounts y_i are then combined with unit costs to produce *output measures of volume and price* for non-market production:

$$Q^t_{output} = \frac{\sum_i y_i^t c_i^{t-1}(w^{t-1})}{\sum_i y_i^{t-1} c_i^{t-1}(w^{t-1})} = \frac{C^{t-1}(y^t, w^{t-1})}{C^{t-1}(y^{t-1}, w^{t-1})} \quad P^t_{output} = \frac{\sum_i y_i^t c_i^t(w^t)}{\sum_i y_i^t c_i^{t-1}(w^{t-1})} = \frac{C^t(y^t, w^t)}{C^{t-1}(y^t, w^{t-1})} \tag{1}$$

For simplicity, we have assumed that the technology has constant return to scale so that unit cost $c_i^t = \frac{C_i^t(y_i^t, w^t)}{y_i^t}$ is independent of quantity y_i^t[11]. The quantity index Q^t_{output} measures the cost of producing the vector y^t using technology and input prices of period $t-1$, divided by the actual cost at period $t-1$, meaning both technology and input prices are held constant. The price index P^t_{output} captures variations in unit costs when quantities are fixed. It thus encompasses changes in both input prices and technology. Schreyer (2012), following Diewert (2008), breaks down the price index in those two components:

$$P^t_{output} = \frac{C^t(y^t, w^t)}{C^{t-1}(y^t, w^{t-1})} = \overbrace{\frac{C^t(y^t, w^t)}{C^t(y^t, w^{t-1})}}^{\text{only } w \text{ changes}} \overbrace{\frac{C^t(y^t, w^{t-1})}{C^{t-1}(y^t, w^{t-1})}}^{\text{only technology changes}} \tag{2}$$

The first term $\frac{C^t(y^t,w^t)}{C^t(y^t,w^{t-1})}$ singles out changes in input prices. The second term $\frac{C^t(y^t,w^{t-1})}{C^{t-1}(y^t,w^{t-1})}$ isolates the contribution of changes in the production technology, and equals the growth rate of the inverse total factor productivity. Unlike the usual labor productivity defined as value added divided by hours worked, total factor productivity accounts for the joint productivity of all inputs - labor, capital and intermediate consumption - in the production process. Those concepts of productivity are still comparable for education since labor is its main input.

Both terms of the decomposition cannot be computed as such since $C^t(y^t, w^{t-1})$ is unobserved. However, $\frac{C^t(y^t,w^t)}{C^t(y^t,w^{t-1})}$ is comparable to the input price index P^t_{input} that can easily be derived from the data, though it also captures the change in production cost due to substitution between inputs. Hence, the residual $\frac{P^t_{input}}{P^t_{output}}$ reflects the change in *apparent factor productivity* AFP defined by $\frac{AFP^t}{AFP^{t-1}} = \frac{Q^t_{output}}{Q^t_{input}} = \frac{P^t_{output}}{P^t_{input}}$. The apparent factor productivity AFP^t is close to the total factor productivity usually defined in production function, but it also includes the effect of substitution between inputs (details in appendix A.3). This break down of output price gives us a decomposition of value growth into volume, price of input and efficiency :

$$\frac{Value^t}{Value^{t-1}} = \overbrace{Q^t_{output}}^{\text{volume}} \times \overbrace{P^t_{input}}^{\text{input prices}} \times \overbrace{\frac{1}{AFP^t/AFP^{t-1}}}^{\text{inverse productivity}} \tag{3}$$

When input prices and output volume are held constant, a rise in productivity naturally corresponds to a negative contribution to value growth, as it reduces the total costs of producing a certain quantity

[10] The number of pupil-hours is the sum of the hours of teaching received by all the pupils. The European System of Accounts states that the number of pupils should be used if the number of pupil-hours is not available. For higher education, the number of students is more relevant because a large part of the learning process happens outside of classes.

[11] The production function exhibits constant return to scale at each period but the total factor productivity can evolve over time. For example, if y_i^t is the number of pupils and class size is smaller at t than $t-1$, total factor productivity decreases and $c_i^t(w) > c_i^{t-1}(w)$ for any input price vector w

with a given set of input prices. When there is little substitution between inputs like for the education service, AFP^t is close to the growth rate of total factor productivity.

The assumption stated earlier — that general government produces public services with maximum efficiency — matters not only when an input volume measure is used, but also for the output volume measure, since unit costs are used to weight the strata. If the apportionment of expenses across education levels is suboptimal, volume changes capture less precisely the effect on pupils' intertemporal utility. In France, expenditure per student in primary education is lower that for the other levels, while this is not the case in most OECD countries according to (OECD, 2025)[12]. Cour des comptes (2025) suggests that this allocation may be detrimental to students, which means that the output volume measure could underestimate the effects of a change in the number of pupils in primary education compared relative to other levels. More broadly, the differences in expenditure per pupil across developed countries are more likely to reflect deviations from the optimal allocation of public spending than structural differences. This may bias cross-country comparisons of volume growth and purchasing power parities, since the weights assigned to each level of education are only imperfectly related to the associated welfare gains for the pupils and the population of each country.

3.2.3 Outcome volume measures

Following Schreyer (2012), we define the *outcome* H of a production as a state of the world valued by consumers, such that utility can be written as $U = U(H)$. The outcome $H = H(y, \Omega)$ depends on quantities $y_1, ...y_N$, but also on a range of external factors summarized in a vector Ω. Therefore, the outcome is affected by output, but it is distinct from output, thus it is not the target of national account measurements. Schreyer (2012) underlines that outcome can be seen as an intermediate between consumption of output and utility: households consume good and services to reach an outcome associated to a higher utility.

Defining an outcome variable is challenging because it should be a quantified variable that capture all the benefits of a public service. For education, those could include cognitive, technical, and social skills; future productivity and innovation; well-being of pupils and teachers; social cohesion and reduced crime; as well as political commitment. Schreyer (2010) distinguishes *direct* outcomes like the state of knowledge of the population, and *indirect* outcomes such as employment possibilities, higher wages and GDP growth.

A conventional economic outcome for the education service is human capital, defined as the stock of knowledge and skills that enhance individuals' productive capacity and labor market value. Human capital of pupils is usually proxied by the discounted sum of their future income. Human capital is only one possible outcome of the education service alongside others such as grades and innovations. It has the advantage of being a clearly defined quantitative measure extensively studied in the literature and linked to a standard economic variable (income). It is also strongly correlated with other outcomes.

The concept of human capital follows from the observation that education generates investment returns in the form of future labor income, so the knowledge and skills acquired by an individual can be considered as a type of capital, even though education is typically treated as a consumption service in national accounts. In their seminal article, Jorgenson and Fraumeni (1989) integrate human capital in the US national account, and UNECE (2017) provides a methodology to create a suited thematic account. Human capital is not is the asset boundaries of the SNA because it requires economic ownership, and human capital is embodied in persons who cannot be owned.

However, human capital does not capture important dimensions of the education system such as working conditions in schools, reduction of inequalities, acquisition of social skills and future well-being at work (Sen, 1997). Haveman and Wolfe (1984) identify a wide array of effect of schooling including

[12] In France, the ratio of the cost per pupil in primary education to the average cost per pupil in primary to post-secondary non-tertiary education is 85 %, whereas it is 96 % for OECD countries and 95 % for the EU25 in 2022.

better health, better working conditions and fringe benefits, attainment of desired family size, consumer choice efficiency, crime reduction, and the cognitive development of children of educated parents. Therefore, other outcome variables such as reported well-being of students, life expectancy, criminality rate, IQ test of next generation children or use of contraceptives could serve as complements to human capital measures, capturing dimensions which are not included in this concept. A difficulty is that many educational outcomes — such as the quality of family life and community relationships — remain difficult to capture due to the lack of reliable quantitative measures. Moreover, aggregating multiple outcomes requires expressing them in a common metric, with monetary valuation being the most natural candidate. Haveman and Wolfe (1984) addresses this issue by estimating the private costs associated with these outcomes, under the assumption that marginal value is equal to marginal cost. These aspects are already partly captured by human capital to the extent that they raise future productivity, but they also directly affect the intertemporal utility of pupils, teachers, and society as a whole — effects that are not captured by the concept of human capital. However, the direct effect of production processes on well-being is generally not accounted for in national accounts, as pointed out by Stiglitz et al. (2009), so the limitations of human capital noted here reflect broader blind spots of national accounting principles more generally. In this respect, one could argue that these effects should not be included in order to preserve coherence with the rest of the national accounts, even though some of them — to the extent that they affect intertemporal well-being — could in principle be perceived and monetized by consumers in a market context.

Even an outcome-based measure encompassing all relevant dimensions of education would remain unsuitable as a volume indicator in the national accounts, as it would still be influenced by a wide range of factors Ω that are unrelated to the production of the public service. Human capital, for example, is shaped by the cultural and economic context of upbringing, as well as by migration, labor market conditions, and workplace training. International conventions therefore prohibits the use of outcome volume indicators. In principle, addressing this problem would require identifying, for each year, the causal effect of the education system on outcomes, all other things being equal. Isolating the effect of the education system as a whole is challenging, as it depends on multiple characteristics with long-term impacts, for instance through innovation (Biasi et al., 2021) or the formation of generalized trust (Charron and Rothstein, 2016). However, it is possible to identify the impact of specific key indicators on well-defined outcomes, such as individuals' future wages. Relying on existing estimates from the literature, we can estimate the contribution of these indicators to the evolution of outcome. These estimates will be used to perform quality adjustments in Section 3.3.2, following the approach proposed by Schreyer (2012).

3.3 Quality adjustment for non-market output volume measures

The quantity variable y_i may not fully account for all product attributes, potentially overlooking characteristics of public goods and services that consumers care about, in the sense that they affect the outcome H. *Quality* encapsulates all those characteristics that distinguish varieties of a given product. Quality changes should therefore *in general* be recorded as changes in volume, and not changes in price (Eurostat (2016), section 1.3), though Eurostat (2016) prohibit them for non-market products for comparability reasons as we will see below. Quality adjustments are not specific to non-market production: for market products, nominal aggregates are typically deflated using a constant-quality price index to obtain a proper measure of volume. However, this approach is not feasible for non-market production where volume is measured directly through an output-based method.

In the following paragraphs, we present different ways to account for those quality change, either through the stratification of with an explicit quality variable q_i^t. Besides the conceptual issues outlined below, National Accounts requires some availability and quality constraints (detailed in appendix A.2) on the time series data used as quantities y_i^t and quality q_i^t. In this study, we explore a wide range

of possibilities, and precise to what extent the data leveraged match the criteria required for their implementation in national accounts.

3.3.1 Stratification and implicit quality adjustment

The classification of goods and services into homogeneous products $1 \leq i \leq N$ has an important role in volume measurement. The ESA (paragraph 10.30) indicates that products that have similar unit cost and outcomes for the consumer should be gathered in the same strata. In this respect, each level of quality can be considered as a distinct product. This process is called an *implicit quality adjustment* as the quality does not appear explicitly in the computation of volume indexes, but only in the definition of products themselves. It requires precise financial data, because each quality level should be associated with a unit cost. This approach therefore relies on the same general idea as the input measure of volume (section 3.2.1 above), given that the difference in quality across products is measured using the relative gap in production costs. Consequently, it cannot capture quality changes that have no effect in cost, like a change in teaching practices[13].

Schreyer (2012) shows that if two products that are equivalent for consumers but have different unit costs are considered as distinct in the stratification, price and volume measurements will be biased. Indeed, in this case, a rise in the quantity of a cheaper product will reduce the volume instead of lowering the price index. This underlines the relevance of consumer preferences and thereby the importance of the outcome variable in defining a stratification and applying implicit quality adjustment. It also emphasizes that a finer stratification is not necessarily better.

To ensure comparability across countries, the ESA 2010 (Eurostat, 2016) states that the public service of education should be disaggregated at least at the level of the CPA 2.1, which is based on ISCED 1997 classification of education levels[14]. It also recommends refining the classification of tertiary education for the various ISCED fields of higher education[15] to account for costs heterogeneity. The ESA 2010 mentions that quality can be indirectly captured by a finer stratification, meaning implicit quality adjustment are a priori allowed in the current EU standards. There is no clear boundary between a standard output volume measure recommended by the ESA and an implicit quality adjustment as the basic output method is already stratified. We can consider that a volume measure for education includes an implicit quality adjustment if the stratification include other dimensions than the levels of education defined by ISCED classifications.

3.3.2 Explicit quality adjustment

Implicit quality adjustment of non-market output only captures quality through its effect on cost. It requires detailed financial data and it is better suited for discrete quality measures. To overcome those limits, explicit quality adjustment can be leveraged, meaning that a *quality indicator* q_i explicitly appears in the formula of the volume measure. The variable q_i is a characteristic (or a vector of characteristics) of the product that affects the utility of consumers but is not reflected in the quantity y_i. The ESA 2010 prohibits explicit quality adjustment of non-market output to guarantee

[13] Even when a quality change affect unit cost, the resulting volume can be unsatisfactory. For example, a stratification based on class sizes would imply that an additional pupil in a class would bring half the amount of volume if the class is twice larger, as her unit cost would cost about twice as much. There are reasons to think that this effect of class size on volume is to large, as we will see in the next section.

[14] pre-primary education, primary education, general secondary education, technical and vocational secondary education, post-secondary non-tertiary education, tertiary education, educational support services

[15] 0. General programs; 1. Education; 2. Humanities and Arts; 3. Social sciences, business and law; 4. Science; 5. Engineering, manufacturing and construction; 6. Agriculture; 7. Health and welfare; 8. Services; 9. Other

comparability across countries[16], whereas the ESA 1999[17] and the SNA 2025[18] allows them (table 4 in Appendix summarizes the international conventions). To the best of our knowledge, no country currently incorporates an explicit quality adjustment for non-market services in its GDP calculations [19]. Quality can be proxied by input characteristics such as teacher experience, service characteristics such as class size, or outcome-based measures such as PISA test scores and school climate indicators. Quality should be independent of factors external to the production process, as discussed below.

For an explicit quality adjustment, the issue then is how to combine quantity y_i and quality q_i into a unique volume measure $v_i(y_i, q_i)$. This question constitutes the main contribution of this study, as most of the applied literature on quality-adjusted non-market output relies on methodologies that lack an explicit conceptual framework for addressing this issue. We assume that v is homogeneous of degree 1 in y_i, so that it behaves like the unadjusted volume when quality is constant. Hence, there is a function f such that for any (y_i, q_i), $v_i(y_i, q_i) = y_i f_i(q_i)$. The challenge is to determine which variation in quality is equivalent to a given variation in quantity: for example, what reduction of class size would exactly offset in terms of volume a 1% decrease in the number of pupils. This is pinned down by the elasticity of the adjustment function f. A "strong" adjustment corresponds to a very elastic f, that gives a large weight to variation in quality. In its productivity of public services reports (ONS, 2025), the Office for National Statistics gives an equal weight to quality and quantity by setting $f_i(q_i) = q_i$ where q_i is the result at a standard test. However, we could argue that a 1% increase in this test score is actually not enough to compensate a 1% decrease in headcount, since school should be more about inclusiveness than performance, or because the grades have small effects on pupil future wage and well-being. This underscores the need to establish a theoretical foundation for incorporating quality into the volume measure. In a market context, the weight of quality can be specified using a hedonic regression that estimates consumer preferences. However, since preferences are hard to infer without market prices, alternative procedures are required.

Schreyer (2012) builds a framework to aggregate quality and quantity change into a volume measure of non-market output based on the *marginal contribution of quality to outcome*, which is also suggested in the 2008 SNA[20]. In a market context, households consume goods and service to achieve an outcome linked to higher utility. This outcome depends on the quantity but also on quality, so the outcome H takes the form $H = H(y, q, \Omega)$. We assume that outcome fully captures the effect of the service on utility, which is a strong assumption given that outcomes are often multi-dimensional and hard to quantify as underlined in Section 3.2.3. Denoting $Y_i = (y_i, q_i)$, this hypothesis means that the marginal utility of a product i is given by its marginal contribution to outcome, multiplied by the marginal utility of outcome:

$$\frac{\partial U}{\partial Y_j} = \frac{\partial U}{\partial H}\frac{\partial H}{\partial Y_j} = \frac{\partial U}{\partial H}\left(\frac{\partial H}{\partial y_i} + \frac{\partial H}{\partial q_i}\right)$$

We assume that $\frac{\partial U}{\partial H}$ is constant, so maximizing utility means maximizing outcome using y and q. Quantity and quality substitute similarly in terms of utility and in terms of outcome because the elasticities of substitution are equal: $\alpha = \frac{\partial U/\partial q_i}{\partial U/\partial y_i}\frac{q_i}{y_i} = \frac{\partial H/\partial q_i}{\partial H/\partial y_i}\frac{q_i}{y_i}$. To reflect the utility of consumers as in a market context, the volume measurement $v_i(y_i, q_i) = y_i f(q_i)$ should approximate utility, and thus have the same elasticity of substitution α. The elasticity of substitution of v is the elasticity of f so we

[16]Paragraph 10.30

[17]Paragraph 10.42, details in Handbook on price and volume measures in national accounts 2001, section 3.1.2.2.

[18]Paragraph 18.130 : « In addition, the quantity indicator of each category can be augmented by an explicit quality adjustment factor. One way of identifying explicit quality adjustment factors is by reviewing the effects that the service has on measures of outcome. ».

[19]One exception may be the Office for National Statistics (ONS) which may explore this in the future, see for instance Heys and Taylor (2025).

[20]Paragraph 15.122: "the quantity indicator of each category can be augmented by an explicit quality adjustment factor. One way of identifying explicit quality adjustment factors is by reviewing the effects that the service has on measures of outcome."

define f such that $\frac{f'(q)q}{f(q)} = \frac{\partial v/\partial q_i}{\partial v/\partial y_i}\frac{q_i}{y_i} = \alpha$. This means that if a 1% increase in quality has the same effect as a 2% rise in quantity in terms of outcome, then this equivalence should also be true in terms of the volume measure and thus the elasticity of f should be 2.

Weale (2025) constructs an analogue quality-adjusted volume index for the non-market health sector. For a quality improvement at unchanged cost, his approach yields a change in the volume of output equal to the cost share of the affected treatment multiplied by the log ratio of QALY gains before and after the improvement. However, in his framework, the ratio of QALY gains is itself the quality index, and the conversion into monetary terms runs through the administrative value of a QALY derived from the thresholds used by medical experts' judgment. No such price exists in education. Our elasticity α plays exactly this role but it must be recovered from causal estimates in the empirical literature.

It is important to note that quality adjustment is not specific to non-market production. For market goods and services, nominal aggregates are typically deflated using constant-quality price indices, often constructed through hedonic regression methods. Hedonic regressions, which have become a cornerstone of price index methodology in national accounts (Triplett, 2006), decompose the market price of a good into the implicit prices of its constituent characteristics — such as processing speed for computers. These characteristics are selected precisely because they should capture the dimensions of the product that generate utility for consumers. In this sense, identifying the relevant characteristics in a hedonic regression for market goods is conceptually analogous to specifying the quality indicators that enter an explicit quality adjustment for non-market services: in both cases, the exercise consists in identifying the attributes of a product that matter for consumer welfare and estimating their contribution to value. The key difference is that, for market products, consumer preferences are revealed through observed prices, whereas for non-market services, they must be inferred indirectly — here, through the estimated effect of quality indicators on an outcome variable, namely future earnings. The fact that the outcome materializes only with a substantial delay is nonetheless a genuine shortcoming relative to hedonic methods, which rest on consumers' currently observed willingness to pay.

We argue that our approach does not align with Murray (2010)'s critique of outcome-based quality adjustment[21]. This critique applies squarely to income-based measures derived from the Jorgenson and Fraumeni (1989) framework where discounted lifetime income determines the level of output. It does not apply to our framework, in which the value of education output remains the sum of production costs, strata are weighted by unit costs throughout and future earnings enter only through the elasticity of f — that is, through the rate at which quality substitutes for quantity within a stratum [22].

Furthermore, the use of future economic benefits to value assets is not without precedent in national accounting frameworks. For assets that are rarely exchanged on markets — and for which no observable price is therefore available — international conventions explicitly recommend estimating value as the present discounted value of future returns. The SNA 2025 states that for certain assets, in particular natural resources, the most appropriate valuation method consists in computing the present value of the future economic benefits associated with their use[23]. In this respect, our approach shares the same underlying logic: when market prices are absent or uninformative, the value of a service or asset is approximated by its contribution to future economic outcomes. The distinctive feature of our framework is that it applies this principle not to the stock valuation of an asset, but to the volume

[21] Murray (2010) rejects valuing education based on the future earnings it generates, on the grounds that this would be akin to valuing a car by all subsequent changes in GDP attributable to its purchase.

[22] Murray (2010) also underlines that the relationship between quality and quantity are arbitrary in a quality adjusted volume index. Our contribution is to replace that choice with a condition, $f_i'(q_i)q_i/f_i(q_i) = \alpha$, in which α is disciplined by causal estimates of the effect of quality on outcomes rather than set by convention.

[23] SNA 2025, paragraph 14.28: "For some assets, mainly relating but not necessarily confined to natural resources, the most suitable method for valuation is the present value of future economic benefits associated with the use of the relevant asset." and SEC 2010 paragraph 7.53: "Standing timber is valued by discounting the future proceeds of selling the timber at current prices after deducting the expenses of bringing the timber to maturity, felling, etc."

measurement of a flow of services [24].

We consider human capital as the outcome variable of the education service, as outlined in section 3.2.3. The seminal Atkinson report (Atkinson, 2005) also suggests relying on future income to quality adjust the education service. Assuming that future incomes evolve uniformly across all individuals and all ages, the human capital of pupils can be simply expressed (up to a multiplicative constant) as $H^t(y^t, q^t) = y^t W^t(q^t, \Omega)$ where y^t is the number of pupils and W^t their average future earnings over their professional life. This definition of outcome has the particularity of relying on a variable that national accounts already measure, namely labor productivity.[25] We argue that outcome reasoning is already present in the core framework of National Accounts since ESA 2010 (10.30) requires strata to group products with similar unit costs and similar outcomes for consumers [26].

Even if We use estimates of the effect of education quality on future earnings to calibrate the effect of quality on volume and not as a direct output measure, those same earnings will later be recorded as part of the value added of the firms employing these cohorts — so the accounts would register both a current increase in the volume of education output and, two decades later, a rise in labor productivity as better-trained pupils enter the labor market. We argue that this does not constitute double counting from a national accounts perspective. Future earnings do not measure the level of volume but only weight quality changes within the index, determining the substitution rate between quality and quantity while the money metric remains unit cost. The value of non-market education output consequently remains the sum of production costs, and nominal GDP is identical with and without the adjustment. The two records at issue thus concern different transactions in different periods, and no aggregate in the accounts sums them. The same structure is standard for non-human capital, where a quality improvement raises both the volume of current capital formation and the future output produced with that capital, and it would likewise arise under a conventional hedonic adjustment of a market education service (see Appendix A.4).

We also presume that there are no general equilibrium effects so the future wage of each pupil is independent of the number of pupils i.e. $\frac{\partial W}{\partial y} = 0$. We acknowledge that ignoring general equilibrium effects is a limitation of our framework. Several channels would tend to reduce the aggregate welfare effect of quality improvements relative to our partial equilibrium estimates: queue effects, whereby individual returns partly reflect gains solely in labor market relative position rather than absolute productivity (Spence, 1973; Thurow, 1975), and labor market congestion, whereby a generalized quality improvement increases the supply of skilled labor and compresses the wage premium (Acemoglu and Autor, 2011). However, a countervailing body of evidence suggests that social returns to education quality substantially exceed private returns, through positive externalities that are not captured by individual wage estimates (Hanushek, 2013), even if these macroeconomic effects could be upward biased due to misspecification (Acemoglu et al. (2014)) . The net general equilibrium effect on α is therefore uncertain. Integrating general equilibrium effects would require both a structural model of the labor market that jointly incorporates congestion effects and human capital externalities, and credible identification of their relative magnitudes — a complex but important research agenda. We therefore retain the partial equilibrium elasticities available in the microeconomic literature as our baseline calibration, acknowledging that they capture private returns and may differ from social returns in either direction. Our framework is flexible: any researcher holding a different view of the appropriate general equilibrium elasticity — whether incorporating congestion effects, the AI effect or positive externalities — can directly substitute their preferred value of α. In this sense, our paper provides a

[24] When a company perform R&D for itself, there is no market price for the service so it is valued as the sum of cost, plus a markup based on the ratio of net operating surplus to production for R&D in the same sector (Eurostat, 2014). This is a way to include the future gains of the investment without computing them explicitly.

[25] This is not the case, for instance, for health services when the outcome could be life expectancy, whose effect on current and future value added is far more indirect.

[26] Murray (2010) himself concedes that reallocation gains must be recorded where immediate substitution is possible, since otherwise no productivity change could ever be measured (p.419)

methodological contribution that separates the conceptual framework from its calibration, allowing the two to be updated independently as new evidence accumulates.

Therefore, with no general equilibrium effects, $\alpha = \frac{\partial H/\partial q_i}{\partial H/\partial y_i}\frac{q_i}{y_i} = \frac{\partial W}{\partial q}\frac{q}{W}$, meaning the elasticity of f_i needs to be equal to the elasticity of future wages to quality: $\frac{f'(q_i)q_i}{f(q_i)} = \alpha = \frac{\partial W}{\partial q_i}\frac{q_i}{W}$. This way, the volume changes are measured in "human capital units": if a 1% increase in quality causes a α % in future income, then the volume measure also raises by α %, as if the number of pupils had grown by α %. Gu and Wong (2012) implements a similar method for Canada: they estimate the causal link between result at a standard test and future wages, then they exploit this coefficient to quality-adjust the price index of education production. Besides offering a theoretical justification for the role of quality in volume measurement, this approach also ensures comparability across quantity indexes constructed from different quality indicators, as each quality-adjusted volume is expressed in a common unit of human capital. These quantity indexes can still evolve very differently if they capture distinct channels through which production affects outcome. The shortcomings of the concept of human capital mentioned in section 3.2.3 carry over to this framework. For instance, if the quality variable affects social skills but not future wages, its weights will be zero under this method and the adjustment will have no effect on the volume measure. In this case, the effect of quality on utility is not captured by the outcome variable so the assumption $\frac{\partial U}{\partial Y} = \frac{\partial U}{\partial H}\frac{\partial H}{\partial Y}$ does not hold.

The first challenge to implement this approach is to find a quantitative indicator which captures a large part of the effects of output on outcome, while not being affected by external factors. A trade-off often appears between these two dimensions, which we refer to as *completeness* and *robustness* (a formal definition is proposed in appendix A.5). Class size is robust because it is completely dependent of the organization of the education service, but it only explains a small part of future wages. On the other hand, standardized test performance is more complete, as academic achievement has more impact on future wage, yet it is less robust insofar as it is affected by pupils' socioeconomic backgrounds. A way to build a more complete quality adjustment is to combine several aspects of quality such as class size and teacher experience. If the different indicators of quality are independent, the corresponding adjustment functions can simply be combined (see appendix A.6).

The second challenge is to find reliable causal estimates for the effect of the quality indicator on outcome. An intermediate variable can be particularly useful when direct estimates are unavailable, provided that the variable of interest affect the outcome exclusively through the intermediate variable. For instance, the effect of class size on future income can be assessed by chaining the two effects : first finding the impact of class size on academic performance (grades), and then analyzing how grades influence subsequent wages.

Most often, we only have an estimate of the elasticity of future wage to quality at a specific point q_i, but we lack the heterogeneity of effects required to pin down the functional form of f. The most simple and common type of functions are constant elasticity $f : q \mapsto q^\alpha$, linear $f : q \mapsto 1 + \gamma q$, or constant semi-elasticity $f : q \mapsto e^{\gamma q}$. The type of parameter estimated in the literature does not necessarily correspond to the functional form chosen for f: a semi-elasticity $\gamma = \frac{\partial H/H}{\partial q}$ found in the literature can be exploited to estimate a constant elasticity function $f : q \mapsto q^\alpha$, for example by evaluating the elasticity α at the average quality value used in the estimation[27]. Nevertheless, the elasticity estimate α remains the first order parameter of the adjustment, while the functional form defining how this elasticity should vary with q is only second order and negligible in most cases. Section 7.2 will illustrate the consequences of different functional forms for f in the case of class size adjustment.

This quality-adjusted volume and corresponding productivity measures aim at assessing more precisely the value created by public services, and make them more comparable to their market counterpart, but they are not built as tools for policy making and can lead to incorrect interpretation.

[27]The estimation gives $\frac{\partial H/H}{\partial q}(q_{est}) = \gamma$ with q_{est} the average value of q used in the estimation process. We can deduce the elasticity at this point q_{est} since $\alpha = \frac{\partial H/H}{\partial q/q}(q_{est}) = \frac{\gamma}{q_{est}}$

We show this in Section A.7 by comparing our approach with the marginal value of public funds, which is a cost-benefit ratio specifically designed for policy decisions.

3.3.3 Volume index with explicit quality adjustment

The Laspeyres quantity index Q^t_{output} can be written as the weighted average of the variation rate of quantities $\frac{y_i^t}{y_i^{t-1}}$, with weights equal to the total cost of each product $C_i^{t-1} = c_i^{t-1} y_i^{t-1}$:

$$Q^t_{output} = \frac{\sum_i y_i^t c_i^{t-1}}{\sum_i y_i^{t-1} c_i^{t-1}} = \sum_i \frac{y_i^{t-1} c_i^{t-1}}{\sum_j y_j^{t-1} c_j^{t-1}} \frac{y_i^t}{y_i^{t-1}} = \sum_i \frac{C_i^{t-1}}{C^{t-1}} \frac{y_i^t}{y_i^{t-1}} \tag{4}$$

When applying an explicit quality adjustment, the variation in quantities y_i^t is replaced by the growth in quality-adjusted quantities $v_i(y_i, q_i) = y_i f(q_i)$ in the formulation of Q^t_{ouptut} at the end of equation 4. The quality-adjusted output volume index is thus:

$$Q^t_{adj} = \sum_i \frac{C_i^{t-1}}{C^{t-1}} \frac{y_i^t f_i(q_i^t)}{y_i^{t-1} f_i(q_i^{t-1})} = \frac{\sum_i y_i^t c_i^{t-1} \frac{f_i(q_i^t)}{f_i(q_i^{t-1})}}{\sum_i y_i^{t-1} c_i^{t-1}} \tag{5}$$

The adjusted chained volume is simply the cumulative product of Q^t_{adj}, starting from the base year. Accordingly, the growth rate of chained volume is $Q^t_{adj} - 1$ and the contributions of each product to the variation of the chained volume are

$$contribQ^t_{adj,i} = \frac{C_i^{t-1}}{C^{t-1}} \left[\frac{y_i^t f_i(q_i^t)}{y_i^{t-1} f_i(q_i^{t-1})} - 1 \right] \tag{6}$$

These contributions can be split into a quality and a quantity part:

$$contribQ^t_{adj,i} = \overbrace{\frac{C_i^{t-1}}{C^{t-1}} \left[\frac{y_i^t}{y_i^{t-1}} - 1 \right]}^{\text{quantity contribution}} + \overbrace{\frac{C_i^{t-1}}{C^{t-1}} \frac{y_i^t}{y_i^{t-1}} \left[\frac{f_i(q_i^t)}{f_i(q_i^{t-1})} - 1 \right]}^{\text{quality contribution}} \tag{7}$$

If we assume that the variation in quantity and quality are small, then $\frac{y_i^t}{y_i^{t-1}} \approx 1$ and $\frac{f_i(q_i^t)}{f_i(q_i^{t-1})} - 1 \approx \alpha \frac{q_i^t - q_i^{t-1}}{q_i^{t-1}}$, with α the elasticity of f, which is equal to the elasticity of future wage to quality in the case of education. Hence, at first order, the contribution of quality for product i is simply the product of three terms: the share of the product in total value, the elasticity of outcome to quality, and the relative change in quality:

$$contribQuality_i^t \approx \frac{C_i^{t-1}}{C^{t-1}} \times \alpha \times \frac{q_i^t - q_i^{t-1}}{q_i^{t-1}} \tag{8}$$

This expression, which is only valid for small variations, is a useful way to understand intuitively how quality affects the volume measurement.

The next sections apply different methods presented in the conceptual framework to France's non-market education sector. The analysis focuses on measuring output volume, incorporating explicit quality adjustments based on the sector's contribution to outcomes—a key innovation of this study.. The remainder of the paper illustrates applications of the methods developed in the conceptual framework. Table B in the appendix summarizes all tested methods, estimated volumes, data sources, the estimation period, and the stratification used for the calculations.

4 Input method

The input method estimates the volume of non-market service output using indicators of input quantities. An increase in input costs is therefore recorded as a price change, although it may lead to a volume increase, for example if the quality of output improves. Although the input method does not yield an explicit measure of productivity, computing input-based volume remains essential. First, generating productivity estimates necessitates accurate measurement of input volumes. Second, input-based volume remains an essential benchmark in the context of non-market education. The divergence between output- and input-based measures highlights the sensitivity of volume growth estimates to methodological assumptions, particularly the selection of quantity indicators. Conversely, a strong alignment between these indices suggests that observed changes in volume primarily reflect variations in resource utilization.

As outlined in section 3.2.1, several variants of the input method exist. These methods leverage different data sources and stratifications to construct the same conceptual object, namely the Laspeyres quantity index of inputs denoted Q^t_{input} in Section 3. The standard method is the "input-price" approach which deflates production costs (intermediate consumption, consumption of fixed capital, compensation of employees) using corresponding price indices. This approach is currently applied to collective non-market services such as public administration and defense in the French national accounts [28]. For collective services, input methods are rated at the same recommendation level as output methods by Eurostat (Eurostat, 2016) — both classified as category B — due to the difficulty of identifying a meaningful output measure. This stands in contrast to individual services, for which output-based approaches are preferred and rated as category A.

Alternatively, one can use input quantities and prices to directly compute a Laspeyres quantity index, rather than deflating costs. For labor, this means measuring the evolutions of staff headcounts by category, each weighted by their respective remuneration as in Gemmell et al. (2017). This method accounts for the composition of the workforce by distinguishing different types of teachers and other employees. Section C.3 in Appendix details this method and compares different measures of the labor volume[29].

A third approach, defined as the "market-price" input method, uses the price index of a comparable market service instead of the input-price index (Mitchell et al., 2022) to deflate total cost. This method may be of interest when the price indices used to deflate inputs are poorly measured. Conversely, its relevance hinges on the existence and comparability of a suitable market counterpart. In this context, we rely on the Consumer Price Index for market-based education services in the COICOP classification[30]. In France, this index covers four main components: tuition fees in fully privately financed secondary and tertiary institutions, partial fees in state-contracted private institutions, private tutoring and individual lessons and adult training and language courses. Beyond these market components, the index also captures household out-of-pocket expenditure on non-market education services — such as registration fees and canteen contributions in public schools — which reflects the public-private financing mix of the non-market sector rather than a market price for educational services. This composition raises three distinct concerns. First, for state-contracted private institutions, teacher salaries are publicly funded and household fees only represent a fraction of the true production cost, so price changes in this component may reflect shifts in the public-private cost-sharing policy rather than movements in the actual price of educational services. Second, the household out-of-pocket component

[28] It is important to note that such an index is frequently revised in National Accounts. Our approach may not accurately reflect the actual figures recorded in National Accounts

[29] One could also consider building similar input index for the consumption of fixed capital and intermediate consumption. However, this would raise comparability issues regarding the sectoral coverage, as the non-market education sector data produced by the French National Accounts Department includes higher education, while our data on teaching volumes do not account for this component.

[30] Division 10: "Education" in the COICOP 1999 classification

of non-market education is similarly driven by financing policy decisions rather than by supply costs or user preferences. Third, even for fully privately financed institutions, the service provided differs substantially from public education in terms of pedagogy, student population, and cost structure, limiting the comparability with the non-market sector. For these reasons, this index provides only a very imperfect imperfect proxy for the market counterpart of non-market education.

4.1 Evolution of inputs in the education sector

Figure 12 in the appendix breaks down the cost of education (including higher education) into its components[31]. Compensation of employees (D1) accounts for the largest share of output, with a relatively stable share around 77 %. The shares of intermediate consumption (P2) and consumption of fixed capital (CFC) remain stable over the period, at approximately 13 % and 9 %, respectively. The residual is attributable to the net balance between taxes on production and subsidies[32].

Two salient features characterize the evolution of the volume of compensation of employees,which is the primary component of non-market education output. Between 2016 and 2020, a notable increase is observed in educational support staff, particularly among non-tenured staff (+84 %, see Figure 13b in the appendix). This category includes temporary guidance counselors, educational assistants (AED), and assistants for students with disabilities (AESH). The expansion of non-tenured educational staff is primarily driven by the recruitment of AESH, whose full-time equivalent (FTE) numbers increased by 90 % between 2013 and 2023, as noticed by the Cour des Comptes (2024) in its assessment of the inclusion of pupils with disabilities in regular schools[33]. Also, as shown in Figure 13, between 2016 and 2021, the number of preschool and primary school teachers increased by 1.1 % in the public sector and by 6.3 % in the private contracted sector. This growth can be explained by several factors, such as the recruitment of additional teachers following the extension of compulsory schooling to age three in 2019[34] or the expansion of short-term substitution schemes aimed at systematically covering short-notice absences, which increased the recorded number of teaching staff over the period[35]. The significant rise in the proportion of contract teachers may have implications for teaching quality and, consequently, for the effective volume of educational services, even in the absence of changes in overall headcounts.

Figure 15b in the appendix presents the chained price indices for the components of non-market education output (NAF 85N). The Consumption of Fixed Capital (CFC) price index can be interpreted as a proxy for the cost of non-residential buildings, given that education investments are predominantly directed toward construction-related products (see Table 6). In contrast, intermediate consumption spans a wide range of products, including meat processing and preservation, specialized construction works, administrative activities, and land transport of passengers, as illustrated by the excerpt from the Input–Output Table in the appendix (Appendix Table 5). From 2013 to 2020, the chained price indices for each input category (Figure 15b) rather exhibit similar trajectories, with annual differentials rarely exceeding 0.5 percentage point. Post-2020, however, the indices diverge markedly in response to the inflationary shock affecting the French economy. Both the CFC price index and the intermediate consumption price index surged — rising respectively by approximately 5 and 3 points between 2020 and 2021 — reflecting the broad-based propagation of inflationary pressures across product markets. By comparison, the wage index increases more modestly (by about 0.15 point over the same period), as the index does not account for bonuses, which were the primary mechanism used by the government

[31] The data used to implement the input method are detailed in the Appendix.

[32] Ideally, we would decompose the compensation of employees by level of education. However, this information is not directly available in the central framework of national accounts and would need to be reconstructed from auxiliary sources.

[33] Cour des Comptes, Inclusion of pupils with disabilities in schools, September 2024. The Cour des Comptes is an independent French organization which aims at assessing the effectiveness of public policies

[34] LOI n° 2019-791 du 26 juillet 2019 pour une école de la confiance

[35] Décret n° 2023-732 relatif au remplacement de courte durée dans les établissements d'enseignement du second degré

to offset the erosion of public-sector purchasing power during this period[36]. Section C.3 in Appendix provides more details on the limitations of this wage index.

4.2 Input volume measures

This section documents the evolution of input-based volume measures of non-market education in France considering several methods commonly used in National Accounts for computing input volume. These aggregate trends serve as a benchmark against which the output-based and quality-adjusted measures developed in the following sections will be compared.

Figure 1 displays three volume indices for the education sector: two input volume measures based on input and market prices, and the output volume currently used in the French National Accounts. Here, the scope covers the entire non-market education service, including higher education, as detailed inputs data are only available at this level of aggregation[37]. After 2018, the output volume exhibits a downward trend, primarily driven by demographic factors. In contrast, the two input volume measures have increased almost continuously over the same period. This growth is attributable to rising quantities of intermediate goods and services consumed by the sector. However, 2019 stands as a notable exception: the market-price-based input volume declined, a phenomenon that may be explained by strong price dynamics in education-related activities (classified under COICOP 10). Therefore, a significant portion of the sector's production change was captured by price effects rather than volume.

The sharp rise in input-price–based measure of volume in 2021 and 2022 is due to the fact that the price index used to deflate compensation of employees (D1) does not account for bonuses, which have been the primary instrument for adjusting teachers' wages in the aftermath of the pandemic and the subsequent inflationary shock (details in annex C.3). The evolution of volume under the input-price–based approach therefore largely reflects the limited reliability of the price index employed, which supports the use of an output-based method, as recommended by Eurostat (Eurostat (2016)).

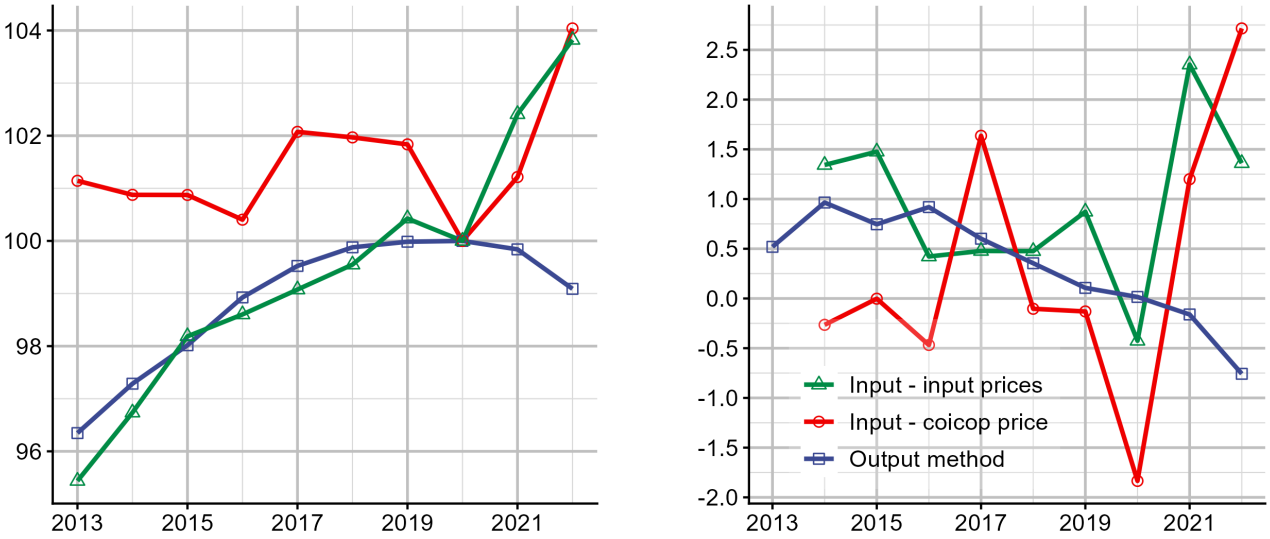
[36]Note that national accounts typically apply additional corrections to account for bonuses and other supplementary wage components when constructing input price indices. Had an input-based measure been used to compute the volume of education, such corrections would likely have been incorporated. Our index does not account for these components, so a more refined index would probably be used in an official national accounts implementation

[37]The thematic account of education provides expenses by level which is required for the output volume measure, but not the breakdown by production factor (intermediate consumption, wage, fixed capital formation). The disaggregation by operations (operating costs, staff, investment) in the thematic account does not correspond to national accounting conventions.

Figure 1: Volume of education production–input and output methods (including higher education)

(a) Chained volume (2020 = 100)

(b) Volume growth rate (%)



Note: The output volume growth rate is computed as the sum of relative change in number of students in each stratum, weighted by the cost of the stratum. The output measure presented here follows the same methodology as the one currently implemented in the French national accounts and relies on the same underlying data, but does not incorporate the specific corrections applied in official statistics. Volume indices based on input methods are computed with input indices, computed as a weighted sum of production costs or as a price index of a comparable market service.
Coverage: Non-market education including higher education
Source: DEPP, Insee, authors' calculations.

Overall, input price indices for non-market education evolved similarly prior to 2020 but diverged sharply following the inflationary shock, with consumption of fixed capital and intermediate consumption prices increasing rapidly. While output-based volume measures show a decline since 2018, largely driven by demographic trends, input-based measures display a continuous increase, reflecting rising use of intermediate inputs and the expansion of non-teaching staff. However, these input-based measures are sensitive to price measurement issues, notably the omission of bonuses in wage deflators, which biases volume estimates upward. Taken together, these findings highlight the limitations of input-based approaches and support the use of output-based measures for non-market individual services, in line with recommendations from Eurostat.

5 Output method without quality adjustment

5.1 Volume in number of pupils

We start with the approach employed in the official French national accounts to quantify non-market education output, using an output volume measure with the number of students as the quantity variable.

Recent and upcoming evolutions We compute the output volume of education services excluding higher education and apprentices in Figure 2[38]. Those levels represents 75 % of total general government

[38]An adjustment for the Covid period was performed in the actual 2020 French national accounts. In 2020 and 2021, a multiplicative factor was applied to reflect the lower number of teaching days. Then, in 2022 and 2023, the volume growth is defined to catch a pre-covid trend, regardless of the number of pupils. This adjustment is not applied here and the standard volume measure is used for every year. Official accounts also distinguish between private and public

spending for education. The strata are pre-primary, primary, lower secondary, special secondary, upper secondary, technical and vocational secondary[39]. Expenses by level are from the thematic account of education[40] (Jeljoul et al., 2016), detailed in Table 8.

Between 2010 and 2018, the volume of education increased with annual growth rates between 0.3 % and 0.7 %. Since 2018, education output volume is decreasing, with the annual decline expected to reach –1.4% by 2029. The demographic downturn affected pre-primary education first, with enrollments beginning to decrease in 2015. Consequently, the output of first degree education has been declining since 2017, whereas the volume of second degree education has only been declining since 2022.

Given that pupil numbers by education level are largely driven by demographic factors, they can be projected with reasonable accuracy over the next decade[41] and used to create volume projections, assuming the structure of unit costs by level is constant at its 2023 value[42]. The upcoming reduction in volume is substantial: total production is expected to be 6 % lower in 2029 than in 2020, and the drop reaches 11 % for first-degree education. The downward trend is expected to persist after 2030, when smaller cohorts completing primary education transition into secondary levels. This trajectory raises important questions for education policies designed to enhance education quality—for example, reductions in class size. Yet, since quality is currently excluded from national accounting frameworks, these policy interventions cannot counteract the projected decline in the measured volume of education services.

education. This distinction has a negligible effect on volume indexes and is not taken into account here.

[39]This stratification is from the thematic account of education, based on ISCED 2011. The level "special second degree" includes classes for middle school pupils in acedemic difficulty (Segpa), specialized schools, regional special education schools (EREA), as well as youths in custodial education. Apprentices are not included in the volume computation as in the French national accounts.

[40]The notion of expense used for the unit corresponds to the activity "Teaching" in the thematic account of education from the ministry of education (DEPP). It does not exactly match the definition of cost used for the value of non-market education in Insee's national account. For example, the expenses of local administration for daycare are accounted in the activity "Teaching" of the thematic account but not in the value of the education service in the central account.

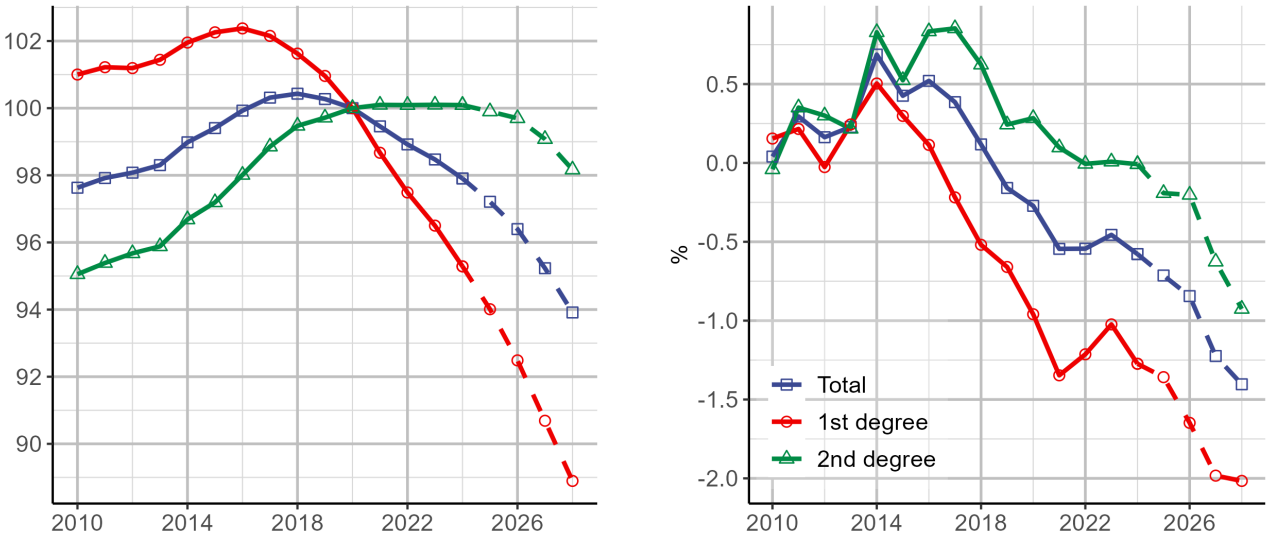
[41]Forecasts for first degree education are from (Croguennec, 2025) and second degree from (Miconnet, 2025)

[42]The definition of levels used for the forecasts does not exactly match that of the education thematic account used to compute the volume index. We therefore assume that the relative differences between those two sources remain constant in the future, in order to produce forecasts of headcounts coherent with the current volume measure. This difference is 13% for upper secondary vocational education and less than 3% for the other strata in 2024. For special secondary education, we only have forecasts for the subset of Segpa (approximately half of the level); we assume that the share of Segpa in special lower secondary education remains constant in the future.

Figure 2: Forecast of education output volume

(a) Chained volume (2020 = 100)

(b) Volume growth rate (%)



Note: Forecasts of the number of pupils from the DEPP are based on demographic previsions.
Coverage: First and second degree education.
Source: DEPP, authors' calculations.

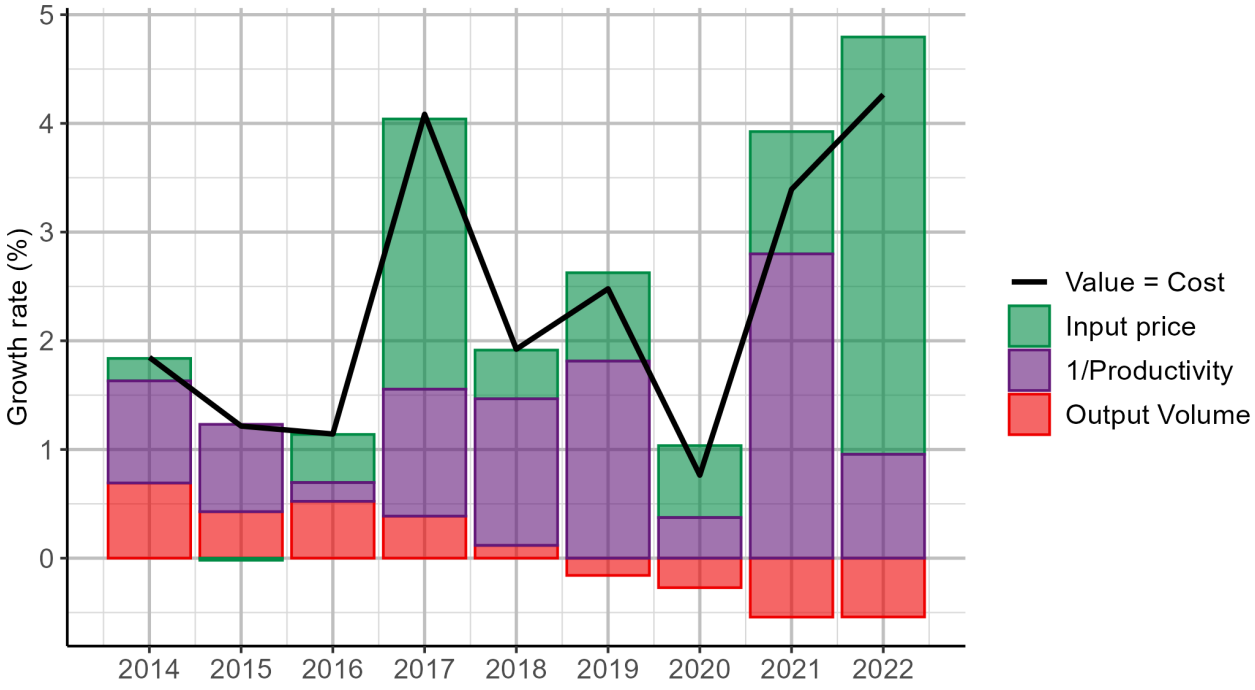
Decomposition into volume, price of input and productivity effects The evolution in the value of production - defined as total cost - can be broken down into the contributions of quantity produced, input prices, and a residual considered as inverse productivity growth, as presented in section 3.2.2.

The value of the education branch in the national accounts is only available for the entire public education sector, including higher education and apprenticeship, so we use the data on expense of the general government (S13) for the teaching activity in first and second degree education from the thematic account of education (Jeljoul et al., 2016)[43]. The index of input price is also available only for the whole education sector, but first and second degree education represent about 75% of the education expense, and we can think that the structure of inputs is broadly similar for the different levels, so we consider that the index of input price is still relevant for the field of first and second degree education only.

Figure 3 shows that volume changes only explain a small part of the variation in value in recent years. The output price of education is driven largely by both input price inflation and productivity changes. Inputs for the production of public education are mostly labor, with wages evolving according to civil servants' seniority and discretionary government measures. The reference wage for civil servant was raised by 1.2% in 2017 and by 3.4% in 2022[44], which explains the two substantial increases in input price visible in Figure 3. Apparent productivity - defined as the ratio of output volume to input volume - has been decreasing every year in the period 2014-2022. This decline can be attributed to an increasing workforce, despite a stable number of pupils. Since class size has no effect on the volume of education measure, this corresponds to a drop in productivity that leads to higher education expense all else equal. However, as detailed in Annex C.3, the productivity evolutions in 2021 and 2022 is biased because the wage index used for non-market education does not include a part of the pay rise, thus the change in input prices and productivity are both underestimated.

[43]Those data are those use to compute the cost per pupil to weight the different strata
[44]Source: https://www.fonction-publique.gouv.fr/etre-agent-public/ma-remuneration/connaitre-le-point-dindice

Figure 3: Contributions of volume, input prices, and productivity to the changes of education production value



Note: The input price is the price index of labor, consumption of fixed capital, and intermediate consumption. The inverse productivity is the ratio of input volume to output volume. A positive contribution of inverse productivity indicates a decrease in productivity that raises total cost, all else equal.
Coverage: first and second degree education.
Source: DEPP, Insee, authors' calculations.

5.2 Volume in pupil-hours

The amount of education produced depends not only on the number of students, but also on the number of hours of teaching that they receive. Therefore, the number of student-hours, defined as the total number of teaching hours received by students, constitutes a more accurate measure of the education production than the number of pupils. The ESA 2010 (Eurostat, 2013) recommends using this measure as the quantity of non-market education output y_i, rather than the number of students. It is already used in many countries according to Schreyer (2010).

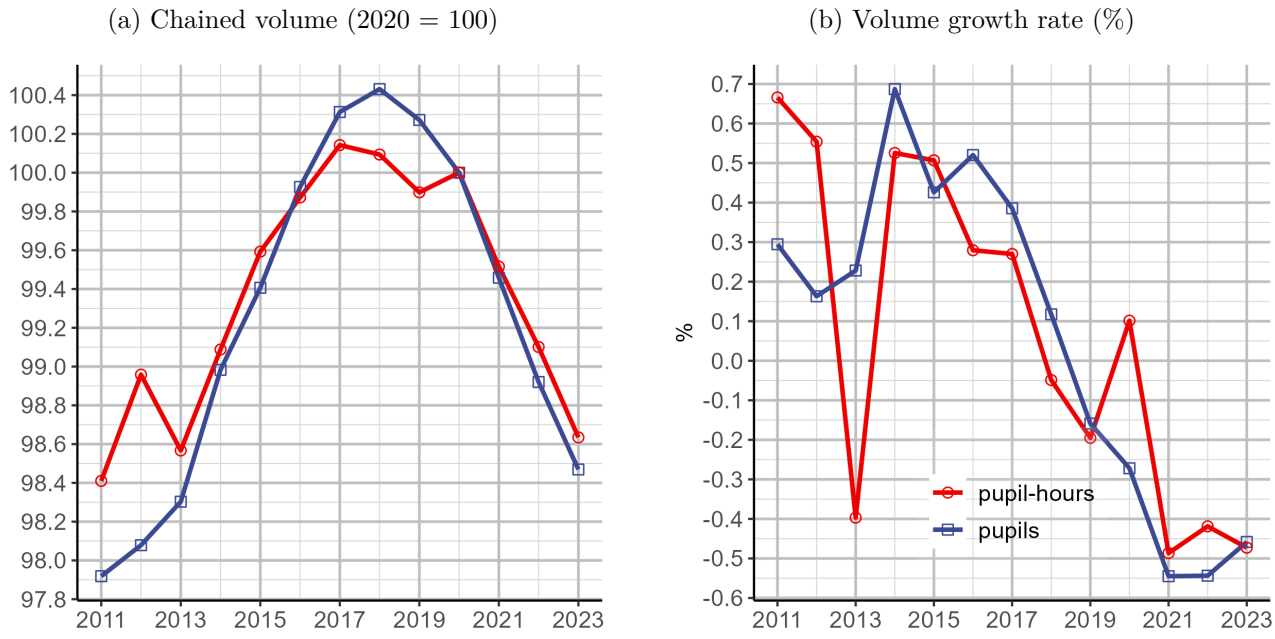
The number of school days per year has been stable since 2000, with 16 weeks of holidays. It fluctuates slightly[45] depending on the timing of public holidays and leap years, but national accounts aim at removing the effect of such variations from aggregates. We therefore assume that changes in the number of hours per student are driven solely by changes in weekly hours.

For pre-primary and primary education, the number of teaching hours per week is identical for all students and has undergone several reforms. The school week was downsized from 30 to 27 hours in 1969, then to 26 hours in 1990, and finally to 24 hours in 2008. For lower and upper secondary education, the number of hours per week of a student depends on her program and choice of courses[46].

[45] From 112 to 117 school days according to https://blog.juliendelmas.fr/?historique-des-vacances-et-du-temps-scolaires

[46] We use data from the H/E and E/S indicators. The H/E indicator is defined as the number of weekly teaching hours divided by the total number of students. The E/S indicator measures the average number of students taught by a teacher during one hour of instruction. For each educational level, we multiply H/E by E/S to obtain the number of hours of instruction attended by a student per week. H/E data for lower secondary education include Segpa programs. We assume that this indicator remains representative of our lower secondary stratum, as Segpa students account for only a small share of enrollment. Similarly, the H/E indicator for general and technological upper secondary education includes post-secondary tracks within upper secondary schools (BTS and preparatory classes). As no specific data are

Figure 4: Volume of education production in number of students and student-hours



Note: The number of pupil-hours is the sum of the number of teaching hours received over all pupils.
Coverage: First and second degree education.
Source: DEPP, authors' calculations.

Weekly instructional hours in secondary education have been broadly stable since 2009. They are higher in vocational upper secondary education (34 hours) than in general and technological upper secondary education (32 hours), and lower at the lower secondary level (29 hours). A peak is observed in both general and vocational upper secondary education in 2011, following the Chatel reform, which introduced two hours of individualized support per week and strengthened specialization tracks. This peak was immediately followed by a decline in 2012 and 2013.

Since 2011, the chained volume measured in student-hours has broadly followed the same trends as the volume measured as the number of students (Figure 4). The peak in weekly instructional hours in upper secondary education in 2011 led to a temporary divergence, after which both series evolve similarly. The small peak in pupil-hours in 2020 is likely due to the uncertainty in the data used as we are looking at small absolute variations[47]. The major shock to the student-hours measure is the reduction from 26 to 24 weekly hours in first degree education in 2008, which caused a 5.5% drop in the volume of first degree education measured in student-hours, whereas annual variations are typically on the order of ±1%.

The student-hours measure does not include any quality adjustment: it does not account for new dimensions of the service provided, but it refines the current quantity measure. It is already permitted - and even recommended - by international accounting conventions. The data are robust and comparable across time, but they are not available prior to 2011 for the second degree.

Unprovided teaching hours The amount of service provided depends on the number of *effective* teaching hours, whereas the volume in pupil-hours volume presented in this section is theoretical, as it does not take into account unprovided teaching hours. The *Tenae* survey measures such unprovided hours for public secondary education institutions since the 2017/2018 school year, but data are publicly

available for special secondary education, we assume a constant weekly volume of 30 hours, close to the lower secondary average. Source: RERS 2024, Table 9.12.

[47]For example, the E/S indicator in vocational high school moving from 15.7 to 15.8 in 2020 represents a quite significant 0.6 % increase for this strata while this variable has only one decimal place

available only for 2022/2023 and 2023/2024 (Stefanou, 2025). Hours are unprovided either because the teacher is absent and not replaced (78% of unprovided hours in 2023/2024), or because the school is closed (22%). The volume of unprovided hours is substantial and exhibits significant variation: it went from 11% in 2022/2023 to 9% in 2023/2024. As a result, the volume measured in effective instructional hours would display a different profile than the one shown in Figure 4, because a change from 11% to 9% of unprovided hours corresponds to an increase of $91/89 - 1 = 2\%$, whereas typical volume changes are in the $\pm 1\%$ range. If we apply this share of unprovided hours to all secondary education strata[48], the volume growth rate for secondary education between 2022 and 2023 becomes +2.1% instead of -0.1%. Therefore, this effective teaching hours measure could lead to significant change in volume indexes and a higher variability over time.

Moreover, simply subtracting unprovided hours from theoretical instructional time to compute a quantity index could underestimate the associated welfare loss, because canceled teaching hours are often unexpected, so they induce delay in the program and deteriorate the overall quality of education received. In addition, they may generate practical constraints, for example for parents who must stay at home to take care of their children.

6 Output method with implicit quality adjustment for pupils with disability

Motivations Enrollment of children with disabilities in regular schools[49] has more than tripled in the last twenty years after a 2005 law for inclusivity[50], from 151,000 in 2005 to 496,000 in 2024[51]. The cost of a pupil with disabilities is higher than the average, and this gap has widened over the period: it was 22% higher in 2006 and 74% in 2026.

Pupils with disabilities receive a specific service of education, as they are often helped by an assistant and offered adapted materials. The category "special second degree" partially accounts for those specificities, as it encompasses disabled secondary students attending specialized institutions, such as localized educational support units (ULIS). However, since 2005, most of additional pupils with disabilities have been integrated into regular classes. Moreover, first degree pupils attending specialized institutions like ULIS are not separated in a distinct stratum in the national accounts. As a result, the costs per education level cover expenses for disabled children, but the volume index ignores the specific services offered to those pupils. Because the cost per pupil with disabilities is higher, an increase in their share is recorded as a price effect unless they are classified into a separate stratum that assigns them a higher weight per student.

Alternative stratification design To address this issue, we test a stratification where each level - except special second degree - is divided into a disabled and a non-disabled group (details in Table 7 in appendix). Then, we reallocate both the headcounts and the costs to those groups based on available information[52]. Total expense for pupils with disabilities are from the Budget reports (program 230,

[48] We use the value for school year *n–1/n* to represent calendar year *n–1*, as data are available for only two school years.

[49] Pupils with disability can be either in a regular environment or in a specialized environment, which covers hospitals and medico-welfare establishments managed by the Ministry of health. Regular schools include *regular classes* (including Segpa and EREA), and *special classes* which can be in localized educational support units (ULIS) or in teaching units for autism (UEMA and UEEA).

[50] Loi n° 2005-102 du 11 février 2005 pour l'égalité des droits et des chances, la participation et la citoyenneté des personnes handicapées

[51] Number of pupils with disabilities by level are from the RERS 2025

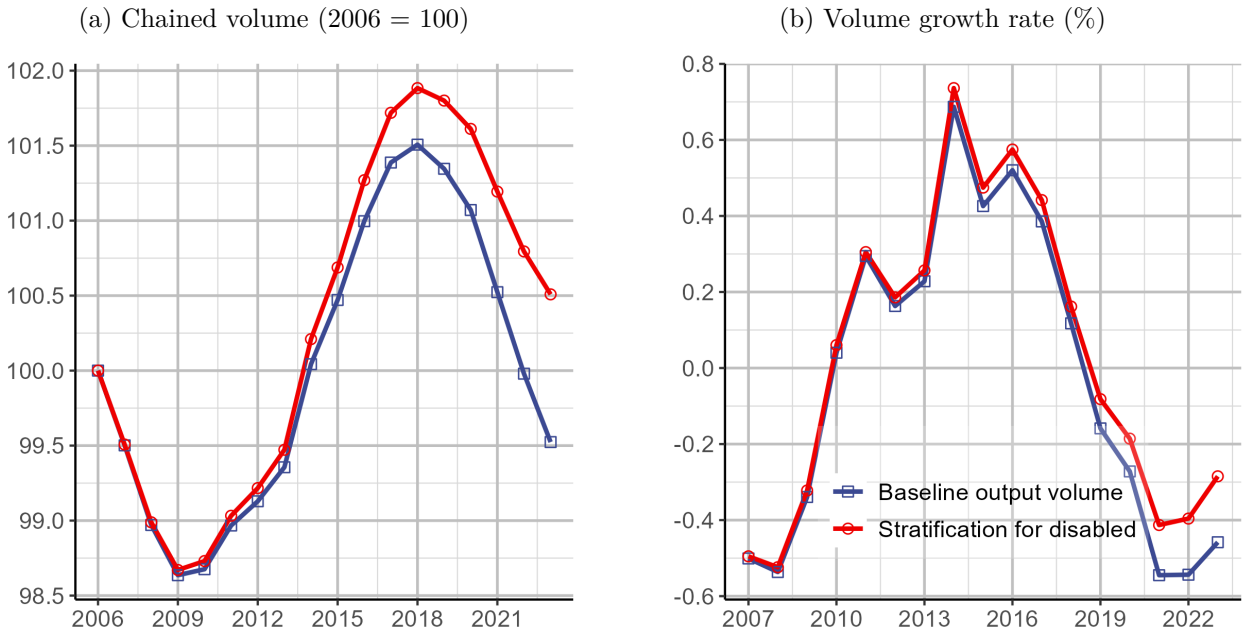
[52] We use the share of second degree pupils in disability across the different second degree levels using data from school year 2024/2025, assuming that the distribution is constant. For first degree, we only have the distribution of pupils with disability by age, so we consider that those aged 2 to 5 are in pre-primary education and the others are in primary education.

action 3[53]). Cour des comptes (2024) underlines that those budget data do not include the wages of specialized teachers, which could raise the expense by 61% in 2022. Therefore, our procedure probably underestimates the unit cost associated to pupils with disabilities, and thereby underrates the increase in volume due to their growing enrollment.

This methodology can be seen as an implicit quality adjustment with quality being the ability of the education system to adapt to pupils with disabilities. As underlined in Section 3.3.1, this methodology the weight of quality is based on costs, which could poorly reflect consumers' preference. Parents of pupils with disability could be willing to pay more than the government for the service of education provided to their children, which would mean that the adjustment underestimates the effect on volume.

Adjusted volume The change in stratification does not affect the total cost of the education service as it only reallocates expenses, but it gives more weights to the increasing number of pupils with disabilities in the volume measure. The cumulated growth in output volume for education from 2006 to 2023 is therefore one percentage point higher with this categorization, from -0.51 % to +0.48 %, as the volume growth rate is slightly higher every year and substantially higher from 2021 to 2023. It is allowed by the EU conventions (Eurostat, 2016), as it simply consists in using a more disaggregated classification of the education services.

Figure 5: Volume of education production using a stratification adapted to disabled children in ordinary classes



Note: Volume index are computed as the sum of relative change in number of students in each stratum, weighted by the cost of the stratum. The stratification suited for pupils with disability is more detailed as it distinguishes pupils with disability from the rest of students in each level of education.
Coverage: First and second degree education.
Source: DEPP, authors' calculations.

[53]We use the payment credits from the annual performance report (RAP): https://www.budget.gouv.fr/documentation/documents-budgetaires-lois/exercice-2023-anterieurs. We allocate the cost of "Action 03" across education levels and between special and regular classes based on the number of students with disabilities, assuming that the additional cost of a student with disabilities is the same regardless of the type of school and the student's individual conditions (individual assistant or not, partial or full-time schooling, etc.). This procedure likely overestimates the cost of schooling in regular classes, since spending per student with disabilities is probably lower in regular classes than in special classes such as ULIS. The cost of non-disabled students in regular classes is obtained by deducting the estimated disability-specific cost from the total cost of each level in order to avoid double counting

The share of students with disabilities in regular classes has tripled, from 0.8% in 2006 to 2.9% in 2023. This increase has been accompanied by a rise in material and human support resources, but it may still have affected the quality of teaching: a teacher who has to devote time to a student with disabilities who does not have an assistant has less time for the rest of the class (Tenne et al., 2023). Teachers can also experience a deterioration in their working conditions, as they are not necessarily trained for this type of task. Consequently, we can think that the increase in pupils with disabilities in regular classes can negatively affect the quality of education for all the students. It is also possible that the inclusion of pupils with disabilities leads to higher tolerance and cooperation for the rest of students. These spillover effects are not accounted for in our estimate.

7 Output method with explicit quality adjustment

Quality is a multidimensional concept that can be examined through various dimensions, depending on the functions attributed to educational services. These include, for example, the system's capacity to transmit civic values and to provide childcare. In this paper, however, quality adjustment is defined in relation to human capital, broadly conceived as the discounted sum of future earnings. Indicators are thus considered based on their impact on future income.

In this section, we present the main quality indicators that we tested: international tests, class size, and the share of novice teachers. Class size and the proportion of novice teachers relate to the inputs and characteristics of educational services, whereas international test scores are typically viewed as an outcome of education. In this analysis, however, we treat them as a quality indicator, reflecting the impact of education on our primary outcome of interest—human capital—which is more closely aligned with the intertemporal utility of students.

Other variables were considered but not retained due to limitations in data comparability or availability, including grade repetition rates, national exam pass rates, inspection reports, teacher wages, teacher qualifications and school climate. Several factors motivate the exclusion of these variables. Changes in grade repetition rates largely reflect policy shifts rather than underlying educational performance, while national exam pass rates are subject to frequent reforms that undermine their comparability over time. Inspection reports, moreover, are predominantly qualitative in nature. In addition, teacher wages cannot serve as volume-based quality indicators without violating National Accounts conventions as they are usually considered as prices. Consistent data on teacher qualifications are also unavailable. Further details on these indicators are provided in Section F in appendix. Although school climate is not included in the main results due to current data limitations, a preliminary methodology has been developed using the available data, leading to negligible effects on volume growth rates (below 0.1 percentage points). This adjustment is presented in Section G in appendix.

7.1 International tests

Student achievement on standardized tests is the most commonly used measure of the quality of educational services (Schreyer, 2010). These tests directly assess the acquisition of academic skills, which constitutes one of the primary objectives for an educational system. However, the use of student achievement presents several challenges. First, student tests rarely cover all educational levels, thus they only partly capture the cumulative effect of the education system and its potential heterogeneity. These tests do not allow tracking student performance across all grade levels, which would be necessary to isolate the effect of a specific educational level. Second, tests reflect factors which are independent of educational quality, such as socioeconomic and cultural determinants, although these factors are less likely to drive changes over time. The literature has extensively documented how economic and social background affects academic achievement (see, for example, Hanushek et al. (2023) for a review).

We consider the economic returns of student test achievement to adjust the output of education

services. The estimation of returns to human capital, as measured by student test scores, has been the subject of extensive research (Card and Krueger, 1992; Card, 2001; Duflo, 2001; Heckman et al., 2006). Nearly two decades ago, Hanushek and Wößmann (2007) identified hundreds of studies on this topic. Early work primarily used years of schooling as a proxy for human capital. Recent studies across 160 countries estimate that an additional year of education increases annual earnings by an average of 10% (Psacharopoulos and Patrinos, 2018; Montenegro and Patrinos, 2023). However, a growing body of literature advocates for supplementing years of schooling with standardized test scores as a measure of education quality. Three seminal studies pioneered this approach (Mulligan, 1999; Murnane et al., 2000; Lazear, 2003). These studies rely on representative national evaluations that track students as they enter the labor market, showing that a one-standard-deviation increase in mathematics scores translates to an average 12% increase in future earnings. Subsequent U.S. studies confirm this finding (Heckman et al., 2006; Chetty et al., 2011; Kline and Walters, 2016). In France, few studies successfully link student test outcomes to future wages. Recently, Fajeau et al. (2025a) used data from a panel of lower-secondary students (*Panel d'élèves du second degré* (1995)) to estimate the impact of 6th-grade evaluation scores on early-career earnings. They found that a one-standard-deviation increase in 6th-grade scores leads to a 9.5% increase in entry-level wages.

From a partial equilibrium perspective, several arguments suggest that microeconomic estimates may understate the lifetime returns to human capital. First, most studies observe individuals over a limited time horizon, typically around age 25. Altonji and Pierret (2001) show that returns to cognitive skills tend to increase with labor market experience, as employers progressively learn to identify and reward underlying productivity — implying that early-career wage estimates provide a lower bound on lifetime returns. Second, skill-biased technological change may have amplified the demand for skilled labor over recent decades (Katz and Murphy, 1992), though the direction of this effect is uncertain in the context of recent advances in artificial intelligence, which may simultaneously substitute for some cognitive tasks while complementing others (Acemoglu and Autor, 2011).

At the macroeconomic level, cross-country evidence documents a strong association between cognitive skills and long-run GDP growth (Hanushek and Woessmann, 2012), with implied social returns substantially exceeding microeconomic estimates of private returns — consistent with the existence of positive externalities through innovation, generalized trust, and knowledge spillovers. However, this cross-country evidence is subject to well-known identification concerns, including reverse causality and the difficulty of isolating the causal effect of education quality from correlated institutional determinants of growth (Bils and Klenow, 2000; Acemoglu et al., 2014), so it should be interpreted with caution.

A further source of uncertainty concerns the stability of the relationship between measured cognitive skills and future earnings[54]. Recent evidence suggests that returns to cognitive skills as measured by standardized tests have declined since 2000 in several countries including the United States, Sweden and Finland (Deming and Silliman, 2025), even as the demand for college-educated labor continued to grow — suggesting that what is rewarded in the labor market has shifted away from the type of cognitive competencies captured by test scores. This is reinforced by evidence that early-childhood programs can raise adult labor market outcomes without producing lasting gains in measured cognitive skills (Heckman et al., 2006), and that test-score gains from educational interventions tend to fade out over time even when earnings gains do not (Deming, 2009; Chetty et al., 2011). Taken together, these findings suggest that the non cognitive skills most valued by employers are imperfectly captured by standardized assessments [55], which implies that anchoring α to the estimated effect of test scores on wages may systematically understate the true welfare value of quality changes operating through

[54]We are highly indebted to Abe Dunn for this remark.

[55]Standardized assessments of non-cognitive skills do exist. PISA includes a focus on non-cognitive skills in certain editions, such as creative thinking in 2022. However, these data points are not available across the entire time series, making them unsuitable for longitudinal analysis.

non-cognitive channels.[56]

Finally these considerations suggest that the elasticities we borrow from the microeconomic literature are surrounded by substantial uncertainty. General equilibrium effects operate in competing directions: on one hand, the career dynamics identified by Altonji and Pierret (2001) and positive human capital externalities would push the true social elasticity above our partial equilibrium estimates; on the other hand, labor market congestion, positional effects, the declining returns to cognitive skills documented by **?**, and the potential substitution effect of AI would push it below. The net direction of this bias is uncertain, which is why we treat our elasticity estimates as plausible benchmarks rather than precise welfare parameters, and present results across a range of calibrations.[57].

Quality adjustment - Middle schools We model the quality-adjusted volume of education as $v(y_i^t, q_i^t) = y_i^t e^{\gamma q_i^t}$ where y_i^t denotes the number of students, q_i^t is a measure of students' performance, and $\gamma \geq 0$ is a parameter. This specification reflects the empirical evidence that returns to students' performance are characterized by a constant semi-elasticity—i.e., a linear effect in test score levels—rather than a constant elasticity. We consider the average PISA score in reading, mathematics, and science for France as a proxy of student achievement q_i^t. Indeed, PISA survey has several advantages compared to other available national and international tests[58]. PISA survey stands out due to its standardized design, targeting 15-year-old students across 38 OECD member countries and several partner economies, with data spanning from 2000 to 2022. We restrict our analysis to PISA, as it is the data source with the longest observation period[59]. In France, the PISA sample primarily includes 15-year-old students, most of whom are in the first year of high school. The survey also collects data on students' socioeconomic backgrounds and learning environments. The average OECD score across the three domains is approximately 500, with a standard deviation of 100, meaning two-thirds of students score between 400 and 600[60].

The implementation of a quality adjustment also requires calibrating the parameter γ which measures the economic returns of a variation of student performance. We build on the literature and choose two values, a lower bound $\gamma_{inf} = 0,095$ (Fajeau et al., 2025a) and an upper bound $\gamma_{sup} = 0,30$ (Altonji and Pierret, 2001; Hanushek and Woessmann, 2006). Since PISA reflects mainly skills acquired

[56]The rise of artificial intelligence introduces an additional and fundamentally open-ended source of uncertainty. If AI substitutes for cognitively routine tasks, the labor market premium for the skills measured by standardized tests is likely to compress further, reducing α and attenuating our quality adjustment relative to a benchmark without AI. Conversely, if AI complements human capital by amplifying the productivity of workers who can effectively direct, evaluate and improve AI outputs, the relevant human capital metric shifts toward higher-order competencies that are not currently captured by existing quality indicators. These two scenarios are not mutually exclusive and their relative importance is still an empirical question. What is clear is that the definition of human capital itself may be uncertain, with potentially important consequences for the calibration of quality adjustments.

[57]More broadly, it is important to note that our framework is deliberately flexible: any researcher or statistical office holding a different view of the appropriate elasticity can directly substitute their preferred value of α

[58]The International Association for the Evaluation of Educational Achievement (IEA) administers two other major international assessments: TIMSS and PIRLS. TIMSS, first conducted in 1995, includes two components: one for 4th-grade students and another for 8th-grade students. PIRLS assesses 4th-grade students' reading comprehension. Since 2018, France has also conducted national evaluations, designed by the DEPP. These evaluations cover CP (1st grade), CE1 (2nd grade), CE2 (3rd grade), 4th grade, and 2nde (10th grade). It is worth noting that France's performance have the same trajectory in all international assessments (TIMSS,PIRLS and PISA). However, national evaluations present a different trend: while CP scores have declined over the past two editions, returning to 2017 levels, 6th-grade scores have improved in both French and mathematics (see (Figure 21) and Figure 20 in Appendix. Non-systematic national assessments have also been conducted over a longer period, such as the CEDRE programme,

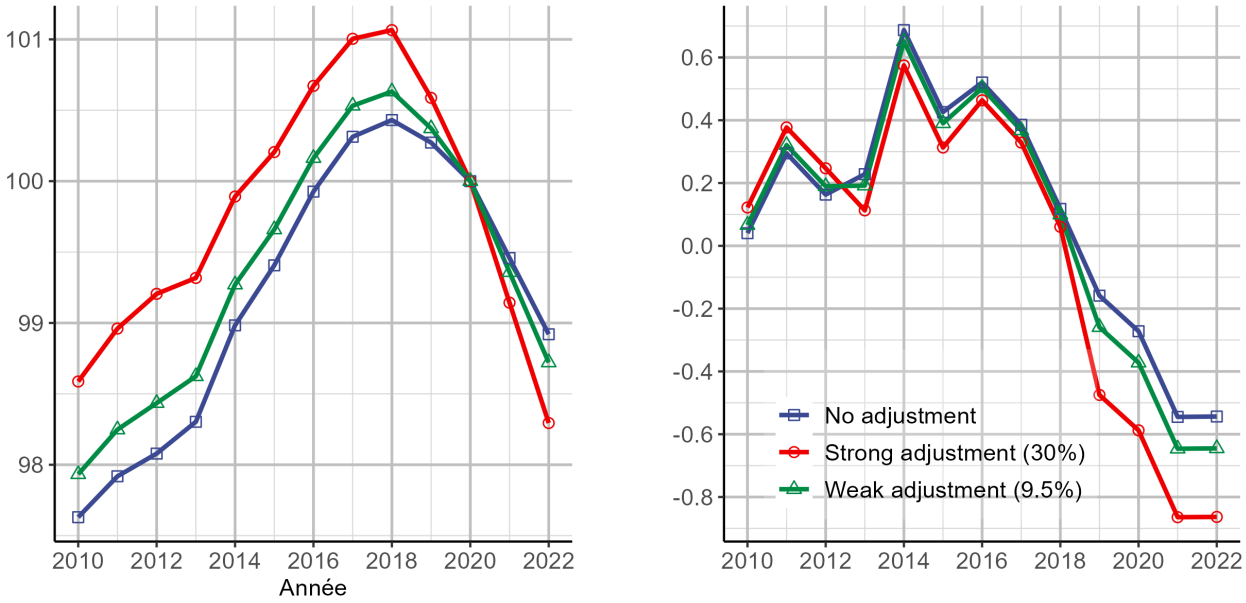
[59]As PISA is conducted every three years, we use linear interpolation for intervening years.

[60]PISA scores are not raw scores, but estimates of latent competencies derived from Item Response Theory (IRT) models. The scale was calibrated separately for each domain during the cycle in which it was first the major domain: reading in PISA 2000, mathematics in PISA 2003, and science in PISA 2006, each time setting the OECD average at 500 points with a standard deviation of 100. Subsequent cycles are reported on these reference scales through linking procedures based on common items, ensuring cross-cycle comparability. In 2015, the transition to computer-based testing led to the adoption of a more general IRT model, but scores were re-equated to the historical scales via linear transformation to maintain comparability. As a result, the OECD average in any given cycle is not necessarily exactly 500 points. For further details, see Okubo (2022) and the PISA technical report 2022 (OECD (2022)).

before high school, we only adjust medium school.

Figure 6: Volume of education production with a quality adjustment for PISA scores

(a) Chained volume (2020 = 100) (b) Volume growth rate (%)



Note: Volume index based on the output method is computed as the sum of relative change in number of students in each stratum, weighted by the cost of the stratum. The adjusted volume is given by $v(y,q) = ye^{\gamma q}$ where q is the PISA score and γ a parameter measuring the economic returns of a variation of student performance. Only lower secondary education is adjusted.
Coverage: First and second degree education.
Source: DEPP, authors' calculations.

A quality adjustment based on PISA scores initially increases the volume at the beginning of the period, reflecting the rise in average performance between PISA 2009 and PISA 2012, driven by an improvement in reading scores from 495 to 505. Toward the end of the period, however, the quality adjustment amplifies the decline in volume. This pattern is consistent with the deterioration in France's PISA performance in the most recent waves. The decline is especially pronounced between PISA 2018 and PISA 2022, with reading and mathematics scores falling from 492 to 473 and from 495 to 473, respectively(see Figure 19 in appendix). However, the difference between the variations of unadjusted volume and quality-adjusted volume remains limited: the difference in growth rates between the unadjusted and strongly adjusted volumes does not exceed 0.4 percentage points. Using the observed PISA mean scores series, annual changes typically range from about 1 to 4 points in absolute value, corresponding to 0.01 to 0.04 standard deviations. With a semi-elasticity of $\gamma = 0.095$ per standard deviation, this implies an effect on wages between 0.1% and 0.4%. Weighting by the share of adjusted strata (approximately 28%) yields an aggregate impact between 0.03 and 0.11 percentage points per year. Using the formula derived in Section 3.3.3, the contribution of the adjustment to volume changes can be decomposed into three factors, namely the share of adjusted strata in total value, the change in average Pisa score, and the semi-elasticity used for future wages, we get for an annual change of 2 points :

$$
\begin{aligned}
&\text{weight of adjusted strata } (\approx 28\% \text{ of total cost}) \\
&\times \text{ change in } q (\approx \pm 2 \text{ points/year} \approx \pm 0.02 \text{ standard deviations}) \\
&\times \text{ semi-elasticity } (9.5\% \text{ wage gain per standard deviation}) \\
&\approx \pm \ 0.05 \text{ percentage points}
\end{aligned}
$$

Quality adjustment - cohort allocation model Each test measures performance at a specific education level. Adjusting only this level may therefore appear insufficient, and a more comprehensive approach would probably be required to extend such adjustments to the National Accounts. Indeed, the test at a given level reflects the stock of human capital whose acquisition depends on prior educational stages. A cohort allocation model can redistribute the share attributable to each educational level for a given human capital measure. It has also the advantage of adjusting several education levels with a single test[61]

Let P denote the stock of human capital measured by PISA among 15-year-olds. Let f_1, f_2, f_3 represent the contributions to P from preschool, primary, and middle school, respectively[62]. We assume the stock-flow relationship is determined as follows:

$$P = k^2 f_1 + k f_2 + f_3 \tag{9}$$

where $k \geq 0$ reflects the relative contribution of earlier educational levels. If $k > 1$, the contribution of earlier levels increases over time. If $k < 1$, earlier levels are less critical [63]. We further assume that the final level (middle school) contributes a fraction $\gamma \in [0, 1]$ of P:

$$f_3 = \gamma P \tag{10}$$

The remaining human capital stock $P(1-\gamma) = k^2 f_1 + k f_2$ is distributed between f_1 and f_2 assuming that the two flows are proportional to their respective weights in the stock-flow relationship[64]:

$$f_1 = \frac{k^2}{k^2 + k}(1-\gamma)P \qquad f_2 = \frac{k}{k^2 + k}(1-\gamma)P \tag{11}$$

$$f_1 = \frac{k^2}{k^2 + k}(1-\gamma)P \qquad f_2 = \frac{k}{k^2 + k}(1-\gamma)P \tag{12}$$

This cohort allocation model reconstructs the flows f_1, f_2, f_3[65] based on three assumptions: i) the functional form of the stock-flow relationship, ii) the relative contribution of educational levels k, and iii) the importance of the final level in the human capital stock γ.

Parameters can be calibrated based on the literature. Note that k captures the extent to which human capital acquired at an earlier educational stage persists and is reflected in outcomes measured at age 15, rather than the marginal return on investment at the time it is made. This is consistent with the well-documented fade-out effect in the early childhood literature, whereby the initial cognitive gains generated by early education programs tend to diminish over time as children progress through subsequent schooling (List and Uchida (2024)). For this reason, we set $k = 0.8$ and $\gamma = 0.2$ thus $f1 = 0.36P, f_2 = 0.44P, f_3 = 0.2P$.

[61] ONS (2025) uses a learning loss indicator that measures the loss of knowledge associated with the failure to achieve a specific grade level. This indicator determines the weights assigned to each educational level within the cohort allocation model. However, this approach relies on the precisely identifying, for each question, the grade in which the corresponding knowledge was required. Consequently, it is not possible to implement the same strategy for France.

[62] While some PISA-tested students are in high school, we assume that their incomplete high school year contributes marginally to their human capital stock.

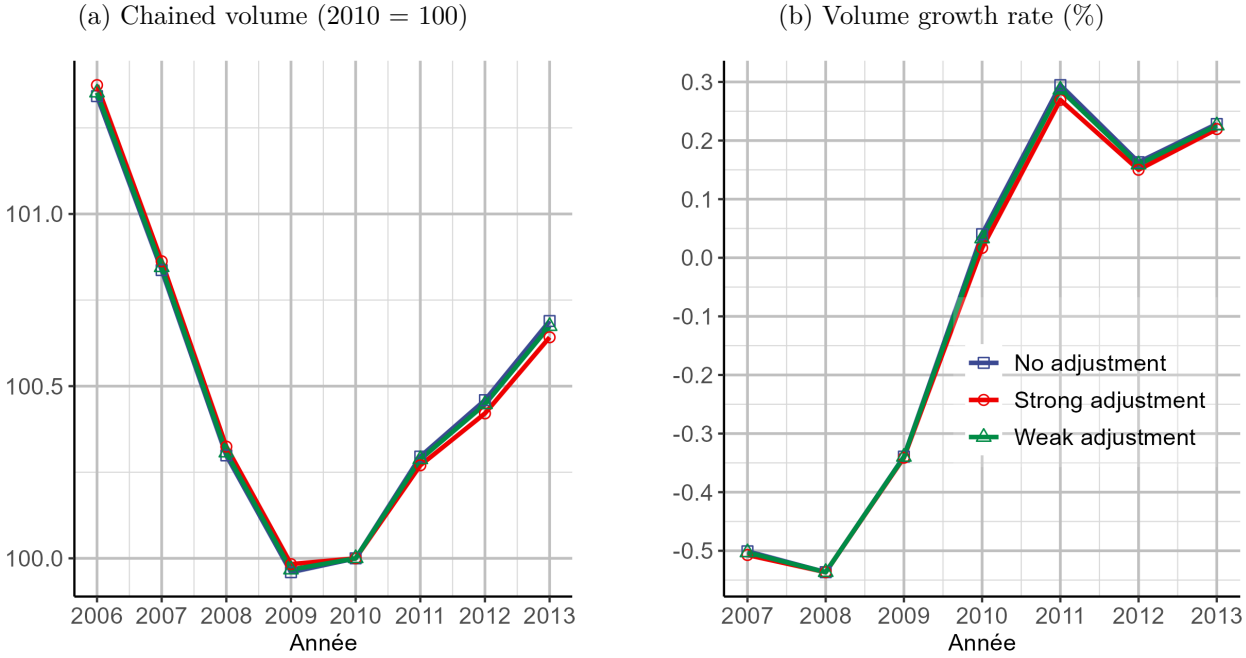
[63] We adopt an additive specification for the stock-flow relationship, which allows the human capital stock to be linearly decomposed into contributions from each educational level. While a multiplicative specification would capture complementarities across levels, it would make calibration substantially more complex. The additive form provides a tractable and transparent first-order approximation.

[64] This additional assumption is required because equation (11) alone does not uniquely identify f_1 and f_2, as it is a single linear equation in two unknowns. We resolve this indeterminacy by imposing $f_1/f_2 = k$, consistent with the idea that the relative weight k captures the relative productivity of each educational stage in generating human capital flows.

[65] Note that f_1, f_2, f_3 correspond to educational levels spanning multiple years: three for preschool, five for primary, and four for middle school. To allocate flows to a specific cohort, the reconstructed flows are uniformly distributed across the relevant years.

We then apply the cohort allocation model to France's average scores in reading, mathematics, and science, replacing the PISA score with the flows from level f_1, f_2, f_3[66], assuming that the effect on future wage is the same for all levels. Since the flows are distributed retroactively, the adjusted education volume is applied with a time lag. We compute the quality-adjusted volume for the period 2006–2013[67]. When no flow is assigned to a given level for a specific year[68], the missing flow is linearly interpolated.

Figure 7: Volume of education production adjusted for PISA scores apportioned across levels using a flow models



Note: Volume index based on the output method is computed as the sum of relative change in number of students in each stratum, weighted by the cost of the stratum. The adjusted volume is given by $v(y,q) = ye^{\theta q}$ where q is the PISA scores apportioned to different levels following a cohort allocation model; and θ a parameter measuring the economic returns of a variation of student performance. The adjusted strata are pre-primary, primary and middle school.
Coverage: First and second degree education.
Source: DEPP, authors' calculations.

Quality adjustments based on flows derived from PISA scores slightly reduce the growth of education volume over the 2006-2013 period. However, the effect is substantially smaller than that obtained when PISA scores are directly used to adjust the middle school stratum. The maximum gap in growth rates is only -0.025 percentage points under the strong adjustment and -0.008 percentage points under the weak adjustment. This limited impact stems from the fact that the most significant variations in PISA scores occur from 2018 onward and are therefore not immediately incorporated into the quality adjustment, as shown in Figure 19. The variations in "flow scores" are then lower than the variation is PISA scores P, partly because they are distributed to several strata according to the cohort-allocation model with a lag.[69]

[66] We still set $k = 0.8$ and $\gamma = 0.2$.

[67] Since the cohort allocation model redistributes the PISA stock of human capital retroactively to earlier educational levels, the resulting flows for primary and pre-primary education in a given year are attributed to cohorts who sat the PISA test several years later. As a result, quality-adjusted volumes for these levels can only be constructed for years for which the corresponding future PISA scores are available. Our PISA sample spans 2006 to 2022, which restricts the quality-adjusted estimates to the 2006–2013 period, since allocating flows up to the most recent levels would require PISA scores beyond 2022.

[68] This occurs, for example, due to the postponement of the 2020 edition because of COVID-19

[69] On the one hand, the share of strata subject to adjustment increases, since we extend the correction to additional education levels with the cohort allocation model. On the other hand, the cohort-allocation model distributes variations

Using the formula derived in Section 3.3.3, we can decompose the contribution of quality ;

$$\begin{aligned}
&\text{weight of adjusted strata (73\%)} \\
&\times \text{change in } q(\approx \pm 0.1 \text{ "flow" points/year} \approx \pm 0.001 \text{ s.d.}) \\
&\times \text{semi-elasticity (9.5\%)} \\
&\approx \pm 0.007 \text{ percentage points}
\end{aligned}$$

7.2 Class size

The relationship between class size and educational quality has been extensively studied, as class size represents a well-defined indicator that is regularly brought to the fore by teachers[70] and policymakers. We define "class size" as the average number of pupils per teacher during a teaching hour, rather than the size of the class seen as an administrative division[71]. In France, first degree class size decreased by 10% in pre-elementary and 12% in elementary between 2016 and 2024 (Figure 22)[72]. This trend is attributable to declining birth rates and policies promoting smaller class sizes in high-priority areas ("REP" and "REP+"). However, under the current methodology—which relies solely on pupil numbers—these policies have no impact on the measured volume of education production, despite their substantial cost. In secondary education, class sizes have remained stable in all grade levels.

Bouguen et al. (2017) and Fajeau et al. (2025c) survey the literature examining the impact of class size on various measures of academic performance. Krueger (1999) is the only study relying on a random control trial, the STAR experiment in Tenessee in 1986, and obtains a significant decrease of performance between 1.8 and 2.7% standard deviation per additional pupil in a class of primary education. Most studies exploit class size thresholds that exist in many countries as an instrumental variable to identify the causal effect of class size. The evidence shows that smaller classes have a significantly positive impact on the academic performance of primary school pupils, between 1.5 et 3 % of standard deviation (Piketty and Valdenaire, 2006; Bressoux et al., 2009; Fredriksson et al., 2012), which is consistent with the findings of Krueger (1999). Nonetheless, some studies leveraging similar strategies like Hoxby (2000) and Angrist et al. (2017) do not obtain a significant effect, and are largely unable to explain this diverging conclusion. Few studies focus on second degree education. Piketty and Valdenaire (2006) find a positive but twice smaller effect of class size in the first cycle compared to primary, and Leuven et al. (2008) observe no significant effect.

The effect of class size is likely heterogeneous: Piketty and Valdenaire (2006) find a twice larger effect on underprivileged pupils, and Connolly and Haeck (2018) show that the marginal effect of an extra pupil is smaller when the class size is already large. The latter is coherent with an intuitive saturation effect for large classes. Those results could justify massive and targeted class size reduction policies, like the one conducted in France in priority education zones.

Long-term impacts of smaller class size are harder to identify in the data. Some studies notice a

in PISA scores across education levels, so that changes in measured quality are mechanically spread across strata. Under the assumption that the impact of test scores on wages is homogeneous across levels, these two mechanisms would, in principle, offset each other, as the same aggregate variation is simply reallocated across a broader set of inputs. However, in practice, we find a smaller overall effect. This is driven by the fact that the relevant PISA variations are measured over a different time window, in which average scores exhibit less movement. This muted variation is further reinforced by the lag structure embedded in the model, which smooths contemporaneous changes in test scores over time.

[70] 97.8% of teachers surveyed for "Exigence de savoirs" want smaller groups (https://www.education.gouv.fr/media/159267/download)

[71] Those differ because in second degree education, a part of teaching hours are with smaller groups, sometimes gathering subsets of different administrative divisions. For second degree education, the average number of pupils in front of a teacher during a course hour is the indicator *E/S* in the French education statistics (DEPP)

[72] Data on class size are from note 25.01 by the DEPP (Evain, 2025) and E/S for secondary education are from note 24.46 (Baradji, 2024).

small positive effect on entering university (Chetty et al., 2011), duration of studies (Browning and Heinesen, 2007) and wage Fredriksson et al. (2012). However, most studies do not find any significant effect on future income (Leuven et al., 2008; Leuven and Løkken, 2020; Chetty et al., 2011; Falch et al., 2017).

Quality adjustment To model how class size q affects the measure of volume, we consider the quality-adjusted volume as the equivalent number of "fully-taught pupils" (FTP) in a class like Collesi et al. (2007). Formally, this can be expressed as:

$$FTP_i(q) = v_i(y = q, q) = q_i f_i(q)$$

Without quality adjustment, $FTP = id$ because each additional pupil increases the volume by one. With a class size adjustment, $dFTP/dq$ is less than one, reflecting the fact that adding an extra pupil increases the headcount by one, but also reduces teaching quality. As a result, the volume of educational output of level i increases by less than one unit. We assume that $dFTP/dq$ is decreasing, meaning FTP is concave, because of the saturation effect shown by Connolly and Haeck (2018). An extreme adjustment would be $f : q \mapsto \frac{1}{q}$, meaning $FTP(q)$ is always equal to one, because the gain in headcount is completely neutralized by the loss in quality[73]. Hence, plausible adjustment function f imply FTP functions intermediate between $q \mapsto q$ and $q \mapsto 1$.

We compare different choices for f to illustrate the ideas developed in section 3.3.2 (specifications in table 1). The adjustments will be applied to both pre-primary and primary, and an adjustment for second degree is tested only as a robustness check, given the limited empirical evidence available on this topic (details in appendix I.1). Our baseline class size adjustment is a constant elasticity function $f^{HC}(q) = q^{-0.04}$, calibrated on the estimate of -0.2 % future income per additional pupil in a class from Fredriksson et al. (2012)[74], corresponding to an elasticity of -0.04 for a standard class size of 22. We compare it to a strong human-capital based adjustment $f^{HC-strong}(q) = q^{-0.18}$ using an upper-bound of the elasticity of -0.18[75], as well as a constant semi-elasticity adjustment function $f^{HC-SE}(q) = e^{-0.002q}$ built on the same estimate from Fredriksson et al. (2012) to analyze the effect of the functional form. We also analyze the specification $f^{Fraumeni} : q \mapsto q^{-0.1}$ from Fraumeni et al. (2008) which has a constant elasticity of -0.1 - intermediate between our baseline and strong human capital adjustments -, and from Collesi et al. (2007) who designs a more complex adjustment function where the class saturates gradually between 10 and 25 pupils, implying that $f^{Collesi}$ is very steep for usual class sizes[76].

Properties of the quality adjustments We analyze the properties of the quality adjustments under the assumption of a single product in the stratification, corresponding to primary education, so

[73] In this case, the volume v_i would be equal to the number of classes and thereby driven by the number of teachers, as for an input volume measure

[74] Fredriksson et al. (2012) is the only study that finds a significant direct effect of class size on income. We can also approximate this effect by chaining two relationships: first, the impact of class size on academic performance, and second, the subsequent effect of grades on future wages, as in Fajeau et al. (2025c). Using the 9.5 % income per standard deviation of Fajeau et al. (2025a) for France, and an average of +2.25% standard deviation when class size decreases by one, we get an effect of $0.0225 \times 0.095 = 0.21$ % coherent with the estimation of Fredriksson et al. (2012)

[75] We combine the 2.8% effect of class size in mathematics from Piketty and Valdenaire (2006) with the 30% effect of grade on future wages of Altonji and Pierret (2001) to obtain an upper bound semi-elasticity to class size $\gamma_{CH-max} = 0.028 \times 0,30 = 0.84$ %, which implies an elasticity $\alpha_{CH-max} = 0.0084 \times 22 = 0.18$

[76] Collesi et al. (2007) defines $FTP^{Collesi}$ as a piecewise continuous function such that every pupil is fully-taught when the class size is below 10 ($FTP(q) = q$ for $q \leq 10$), then the class saturates gradually following a quadratic function until 25 pupils (for $10 \leq q \leq 25$, $FTP^{Collesi}(q)$ is the second degree polynomial such that $FTP^{Collesi}(10) = 10$ and $FTP^{Collesi}(25) = 17$). Above this 25 pupils threshold, the class is considered saturated, so the volume is unchanged by any number of additional pupils (for $q > 25, FTP^{Collesi}(q) = 17$). This makes $FTP^{Collesi}$ strongly concave for usual class sizes, as the slope of FTP^{Colesi} drops from one to zero between 10 and 25. We then define $f^{Collesi} = \frac{FTP^{Collesi}(q)}{q}$ so that $y \cdot f^{Collesi}(q)$ is the number of fully-taught pupils, assuming that classes are all the same size.

Table 1: Adjustment functions

Nom	Definition	Description
No adjustment	$f(q) = 1$	No adjustment
Fraumeni	$f^{Fraumeni}(q) = q^{-0,1}$	Adjustment from Fraumeni et al. (2008)
HC	$f^{HC}(q) = q^{-0.04}$	Human capital adjustment with constant elasticity
HC-SE	$f^{HC-SE}(q) = e^{-0.002 \cdot q}$	Human capital adjustment with constant semi-elasticity
HC-strong	$f^{HC-strong}(q) = q^{-0.18}$	Human capital adjustment using the upper bound estimates of the wage elasticity
Collesi	$f^{Collesi}(q) = \frac{FTP^{Collesi}(q)}{q}$ with FTP a piecewise quadratic function*	Adjustment inspired by Collesi et al. (2007)
HC-2nd	$f^{HC-2nd}(q) = q^{-0.03}$	Human capital adjustment for lower secondary education

Note: The number of fully-trained pupils in Collesi et al. (2007) is a continuous function which increases linearly until 10 pupils, then follows a quadratic concave curve up to 25 pupils and remains constant after that congestion threshold:

$$FTP^{Collesi}(q) = \begin{cases} q & \text{if } q < 10 \\ -1/30 \cdot q^2 + \frac{5}{3} \cdot q - 10/3 & \text{if } q \in [10, 25] \\ FTP(25) = 17,5 & \text{if } q > 25 \end{cases}$$

that the chained volume index simplifies to: $v(y, q) = yf(q)$. Figure 8 plots the FTP and f functions. The number of fully-trained pupils FTP is similar across specifications for small class sizes, but diverges as classes grow larger: a stronger adjustment leads to a larger gap between the number of pupils and fully-trained pupils, and more concave curves. The function f corresponds to the share of fully-trained pupil, i.e. the average volume of a pupil for a given class size q[77]. The steeper is f, the stronger the effect of a change in class size on volume.

Another way to represent the quality adjustment functions is to simulate the effect of a change in q (table 2). Here, we focus on the impact of reducing class size from 22 to 13 students, comparable to what happened in priority education zones in France after the 2017 reform targeting disadvantaged schools. We assume that the total number of pupils is constant to isolate the effect of quality so the effect is simply $\frac{f(13)}{f(22)}$. With the current unadjusted volume index, this measure has no effect on the output volume[78]. The baseline human capital adjustment yields a modest change of 2.1 %, whereas the strong human capital adjustment results in a +9.0 % rise, and $f^{Collesi}$ a +21 % increase as it is strongly decreasing for class sizes between 10 and 25[79]. The choice of functional form has a smaller impact than the choice of calibration, with the constant semi-elasticity form $f^{HC-SE} : q \mapsto e^{\gamma q}$ yielding a +1.6 % change, compared to +2.1 % with the constant elasticity function $f^{HC} : q \mapsto q^{\alpha}$. This is because f^{HC} is more concave, which gives more weight to a reduction of class size if it happens in an already small class, but the difference is second-order as emphasized in section 3.3.2.

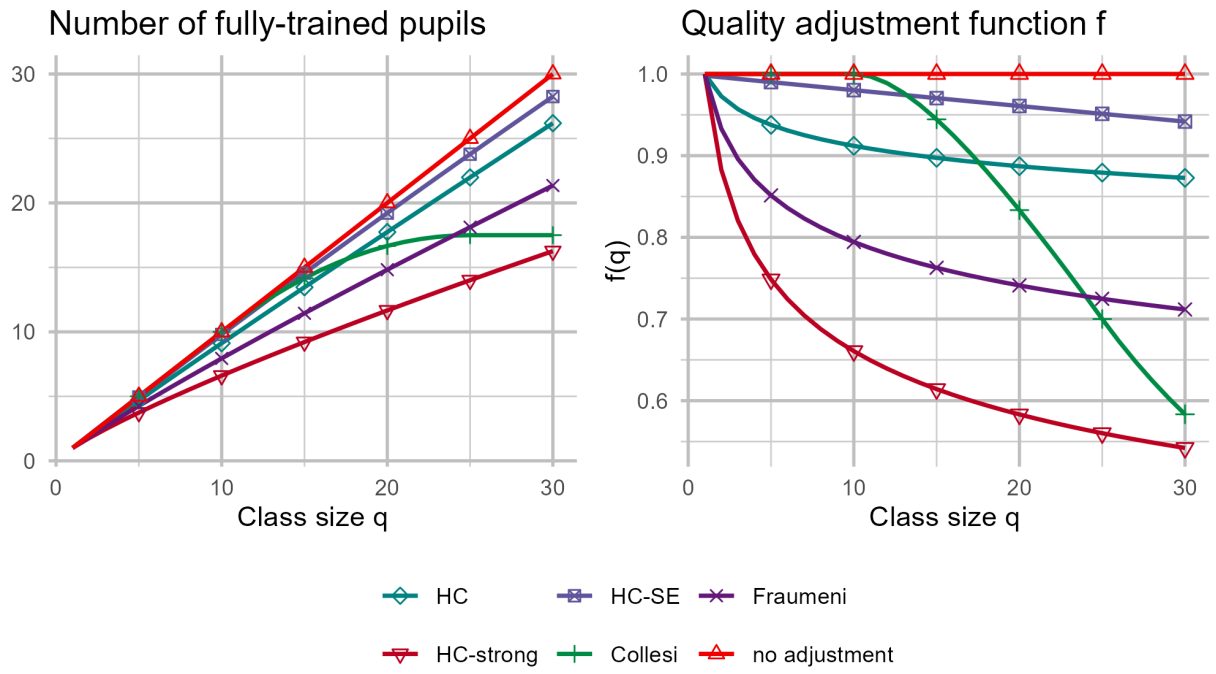
We can also estimate the effect on the volume of input in this simulation to give an idea of the consequence for productivity. Assuming that the only change in input due to class size reduction is

[77] This explains why $f^{Collesi}(q) = \frac{FTP^{Collesi}(q)}{q}$ is not zero for $q > 25$ despite $FTP^{Collesi}$ being flat in this region. The share of fully-trained pupils is not zero for classes larger than 25 because the number of fully-trained pupils is still positive, though the number of fully-trained pupils will not increase when y_t increases.

[78] It could have an effect on the distribution of unit costs across levels in the actual volume measure with several strata, but here we are still considering the volume of the primary education stratum alone.

[79] Since we use only average class size and do not take into account the heterogeneity, the adjustment function f would not be applied to values as low as 13 in the actual computation of adjusted volume because priority education classes are diluted in the pre-primary and primary education strata. The behavior of the adjustment function for very low values therefore becomes particularly relevant if class size heterogeneity is taken into account, and when the adjustment function is very concave

Figure 8: Number of fully trained pupil in a class $FTP(q) = qf(q)$ and quality adjustment function f for different specifications of f



Note: We consider here that there is a single stratum so that the volume is simply the number of fully-trained pupil (FTP) $FTP(q) = qf(q)$, with q the class size. $f^{unadjusted}(q) = 1$ corresponds to the standard volume measure for education relying only on the number of pupils. $f^{HC}(q) = q^{-0.04}$ and $f^{HC-strong}(q) = q^{-0.18}$ are adjustments based on the contribution to human capital with constant elasticity, with $f^{HC-strong}$ using the upper bound of the effect of class size on future wage and f^{HC} using the central estimate. $f^{HC-SE}(q) = e^{-0.002 \cdot q}$ is also based on the contribution to human capital with the same calibration as f^{HC}, but with a constant semi-elasticity functional form. $f^{Fraumeni}(q) = q^{-0,1}$ is the adjustment from Fraumeni et al. (2008). $f^{Collesi}(q) = \frac{FTP^{Collesi}(q)}{q}$ is the adjustment from Collesi et al. (2007), with $FTP^{Collesi}$ a piecewise quadratic function such that $FTP^{Collesi}(q \geq 25) = 17.5$ and $FTP^{Collesi}(q \leq 10) = q$.
Coverage: Pre-primary and primary education

an increase in teaching staff - who represent 60% of teaching costs in 2022 according to the thematic account of education -, the input volume would increase by $0.60 \times (21/13 - 1) = 36.9\%$. This neglects the potential additional classrooms and blackboards necessary for the new classes. The reduction in class size decreases the productivity of education because the relative change in future wage is smaller than the relative change in cost, but this does not mean that it is not a good use of public funds because the absolute gain in welfare of this policy is still higher than the cost according to Fajeau et al. (2025c) which uses similar estimates. They compute a discounted sum of future gross wage at age six of around 400,000 euros per pupil and approximate the cost of halving class size as the annual pay of a teacher, around 53,000 euros. Therefore, the 2.1% volume increase of our central estimate [80] corresponds to a roughly $400,000 \times 2\ \% = 8,000$ euros gains in gross intertemporal wage, which represents $4,800$ euros after tax given a marginal tax rate around 40%. The cost of the measure is only $53,000/21 \approx 2,500$ euros per pupil with 21 pupils per class initially. Hence, the marginal productivity of quality is decreasing, but it still much higher than its marginal cost. In addition, the taxes levied on future higher wage strongly lowers the net cost of this measure, possibly making it self-funded as $8,000 \times 40\ \% = 3,200 > 2,500$, provided the calculation is based on a sufficiently long time horizon (see Section A.7).

[80] Fajeau et al. (2025c) uses a comparable calibration of $11 \times 1.62 \times 0.095 = 1.7$ % gain in future wage, slightly smaller than ours

Table 2: Effect of reducing class size from 21 to 13 students on volume and productivity, keeping the number of pupils constant

	Unadjusted	Quality adjustment				
		Fraumeni	Collesi	HC	HC-SE	HC-strong
Production volume	0.0%	+4.9 %	+20.9 %	+2.1%	+1.6 %	+9.0 %
Productivity ($= \frac{\text{output}}{\text{input}}$)	-36.9 %	-23.3 %	-11.7%	-25.8 %	-21.3 %	-13.9 %
Class size	-38.1 %					

Note: We consider the effect of a class size reduction for a single stratum so that the volume is simply $v(q) = qf(q)$, with q the class size. Different adjustment functions f correspond to different effects of class size on volume. $f^{unadjusted}(q) = 1$ corresponds to the standard measure using the number of pupils. $f^{HC}(q) = q^{-0.04}$ and $f^{HC-strong}(q) = q^{-0.18}$ are adjustments based on the contribution to human capital with constant elasticity, with $f^{HC-strong}$ using the upper bound of the effect of class size on future wage. $f^{HC-SE}(q) = e^{-0.002 \cdot q}$ is also based on the contribution to human capital but with a constant semi-elasticity functional form. $f^{Fraumeni}(q) = q^{-0,1}$ is the adjustment from Fraumeni et al. (2008). $f^{Collesi}(q) = \frac{FTP^{Collesi}(q)}{q}$ is the adjustment from Collesi et al. (2007), with $FTP^{Collesi}$ a piecewise quadratic function. Productivity is an estimate of the effect on apparent factor productivity (volume of output divided by volume of inputs), assuming that only the number of teachings staff is affected proportionally to class size, with teaching staff representing 60% of costs (Jeljoul et al., 2016).
Coverage: Pre-primary and primary education

Adjusted volumes We compute chained volume measures for first and second degree education, adjusting only pre-primary and primary levels (Figure 9). The different adjustments have no effect on volume changes before 2016, since class sizes were relatively stable during this period. Since 2016, class sizes decreased every year, thus the adjustment leads to higher growth rate. The size of the gap with the baseline volume depends on the strength of the adjustment. For instance, Collesi et al. (2007) adjustment is very strong so the volume keeps growing until 2021 despite the decline in pupils headcount, whereas the other adjusted volumes start to decrease in 2019 like the unadjusted volume. This is because the elasticity of the adjustment function for an average class size of 20 pupils is -0.6 for $f^{Collesi}$, 15 times larger than the -0.04 of the baseline f^{HC}, and 3 times higher than the -0.18 of the strong $f^{HC-strong}$.

The maximum gap between the growth rate of unadjusted and adjusted volume is 0.04 percentage point for the baseline f^{HC} and 0.14 for $f^{HC-strong}$ (see Appendix, Figure 24). The impact of the quality adjustment is consequently limited in our contribution to human-capital approach, particularly in light of the fact that standard variations in volume are substantially larger in magnitude, from ± 0.1 % to ± 1 %. Hence, quality accounts for less than a tenth of the volume variation each year. The unadjusted chained volume dropped by 1.53 % between 2020 and 2023 while the baseline adjusted volume decreased by 1.45 % and the strong-adjusted by 1.22 %. Therefore, class sizes have declined, but not sufficiently to offset the decrease in headcounts, as quality accounts for only 5% of the adjusted volume variations between 2020 and 2023 in the baseline specification f^{HC}.

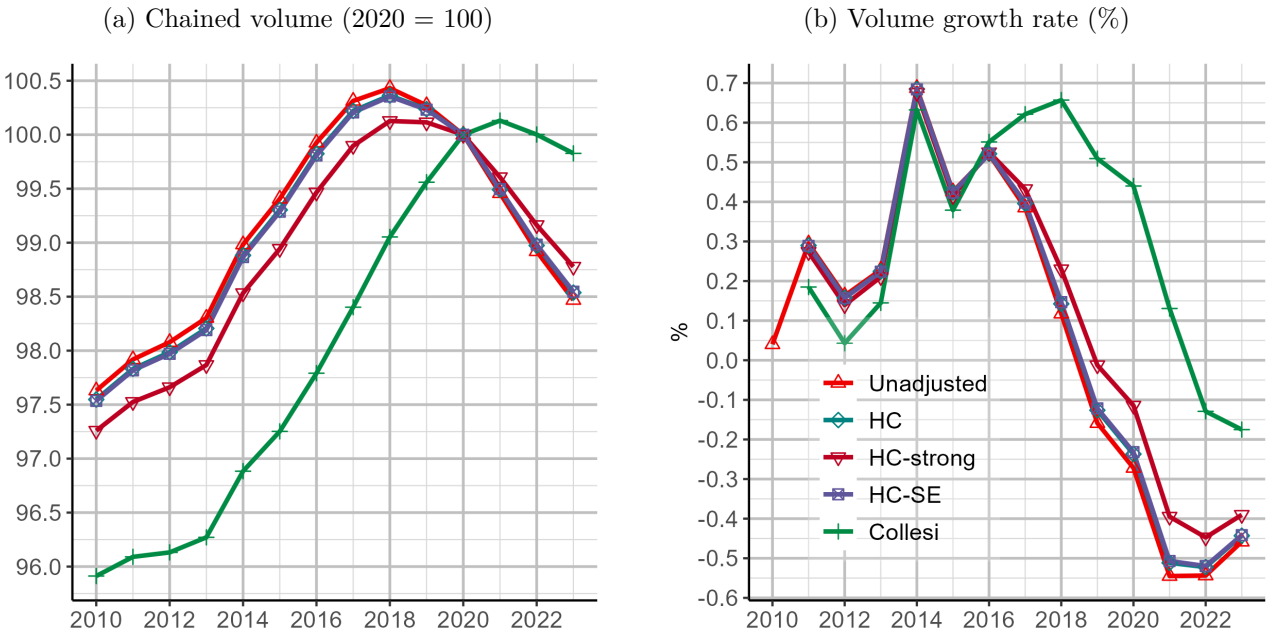
Adjusting the lower secondary education stratum using estimates from Piketty and Valdenaire (2006) also has a negligible effect on the volume of first and second degree education, with yearly contributions always under 0.04 percentage point in absolute value (Figure 24). This is mostly because the corresponding elasticity is small (-0.03 comparable to the -0.04 for first degree), and also because class size in lower secondary have remained generally stable in the last decade with variations below 1% in absolute value for most years and no trend contrary to first degree (Figure 23 in appendix).

The rationale for the small contribution of class size to volume can be illustrated with a simple computation. The yearly relative changes in class sizes are around ± 1 %. The adjusted levels, pre-primary and primary, represent approximately 40% of the total cost of education. The central estimate for the elasticity of future wages to class size is 0.043. Using the formula derived in section 3.3.3, the

effect of the quality adjustment is roughly:

$$\text{weights of adj. levels } (\approx 40\%) \times \text{change in } q \ (\approx \pm 1 \ \%/\text{year}) \times \text{wage elasticity } (0.04) = \pm 0.02 \text{ ppts}$$

Figure 9: Volume of education production quality-adjusted for class size



Note: Only pre-primary and primary levels are adjusted for class size. The adjustments are defined by the adjustments functions $f : q \mapsto f(q)$ where q is class size. $f^{unadjusted}(q) = 1$ corresponds to the standard measure using the number of pupils. $f^{HC}(q) = q^{-0.0425}$ and $f^{HC-strong}(q) = q^{-0.1785}$ are adjustments based on the contribution to human capital with constant elasticity, where $f^{HC-strong}$ uses the upper bound of the effect of class size on future wage and f^{HC} uses the central estimate. $f^{Collesi}(q) = \frac{FTP^{Collesi}(q)}{q}$ is the adjustment from Collesi et al. (2007), with $FTP^{Collesi}$ a piecewise quadratic function such that $FTP^{Collesi}(q \geq 25) = 17.5$ and $FTP^{Collesi}(10) = 10$. The adjustment $f^{Fraumeni}(q) = q^{-0.1}$ from Fraumeni et al. (2008) is not shown because it is very close to the human capital f^{HC}.
Coverage: First and second degree education.
Source: DEPP, authors' calculation.

7.3 Share of novice teachers

Beyond the number of students per teacher, teaching quality depends on teacher-specific characteristics such as experience, motivation, and skill level. Here, the idea is to consider the share of novice teachers as a quality adjustment to account for the negative impact of teacher inexperience on student academic outcomes.

The "teacher effect", i.e. the influence of teacher characteristics on student learning, has been extensively studied. According to Nye and Hedges (2004), who conducted a synthesis of 17 studies carried out in the US between 1971 and 2002, between 7 % and 21 % of variance in student achievement can be attributed to teachers, often exceeding the influence of school-level factors or class size. Using data from the *UTD Texas Schools Project*, Hanushek et al. (2005) show that students taught by teachers with zero years of experience perform significantly worse in mathematics and reading. Accordingly, we consider the share of novice teachers as a quality adjustment in measuring the volume of education.

Data We define novice teachers as teachers with zero years of experience and we estimate their share over the period 2012–2022 using data on the number of candidates admitted to teacher recruitment

exams[81]. For each year and each educational level (primary or secondary), the number of novice teachers is approximated by taking the number of candidates admitted two years earlier and adjusting it for the attrition rate during the training year. The resulting number of novice teachers is then expressed as a share of full-time equivalent teachers.

Between 2012 and 2022, the share of novice teachers averaged 2.9 % in primary education and 3.5 % in secondary education. As shown in Figure 25a in appendix, the share fell sharply in primary education until 2013, reflecting the combined effects of the RGPP policy (which reduced available teaching positions) and the so-called "masterisation" reform in 2010 (which raised degree requirements for candidates). After the end of the RGPP policy in 2012 and the 2013 Peillon reform[82], the share of novice teachers increased in both levels, before declining again after 2017. Overall, between 2012 and 2022, the proportion of early-career teachers grew by about 1.2 points in primary and 0.7 points in secondary education.

Quality adjustment According to Hanushek et al. (2005), a one-percentage-point increase in the share of teachers with zero years of experience is associated with a 0.073 % of standard deviation decrease in students' standardized math test scores. Relying on the estimate from Fajeau et al. (2025a)—namely, that a one–standard-deviation improvement in test scores increases future wages by 9.5 %—this corresponds to a 0.0025 % decline in students' future earnings for every percentage-point rise in the share of early-career teachers. Considering an upper bound of 30 % for the effect of test scores on future wages (Altonji and Pierret, 2001; Hanushek and Woessmann, 2006), the decline would be of 0.0078 %. Then we apply these semi-elasticities $\gamma \in \{0.0025\%; 0.0078\%\}$ with a constant semi-elasticity quality-adjustment function $f(q) = e^{-\gamma q}$, with q the share of novice permanent teachers expressed in percentage.

Results The differences between the growth rates of the quality-adjusted and unadjusted volumes are very small (Figure 26 in appendix), at most 0.017 percentage points for the upper-bound adjustment (0.0078 %) and 0.005 points for the lower-bound adjustment (0.0025 %). These differences mirror the evolution of the share of novice teachers: between 2013 and 2016, when this share increases, the adjusted volume grows more slowly than the unadjusted volume, generating a negative gap; from 2016-2017 onward, as the share declines, the gap becomes positive.

The very small sign of the effect on volume growth rate can be illustrated using the first-order approximation derived in section 3.3.3. The impact of quality on growth rate can be decomposed into three factors, namely the share of adjusted strata in total value (100% in this case), the change in the share of novice teachers (around 0.6 percentage point per year in absolute value, 2 percentage points at most for first and second degrees in 2016), and the very small semi-elasticity of future wages (0.0025 %):

$$\text{weights of adj. strata } (100\%) \times \text{change in } q\ (\approx \pm 0.6\text{ppt/year}) \times \text{semi-elasticity } (0.0025\%) = \pm 0.0015 \text{ ppt}$$

It is clear with this formula that the extremely low amplitude of the effects of quality on volume is due to the tiny semi-elasticity, which reduces the quality effect by a factor of 385 compared to a simple linear calibration. Furthermore, the share of novice teachers fluctuates with no particular trend, so a part of the effects cancel out over time. The average absolute annual contribution of the adjustment to

[81] In France, teachers are primarily recruited through competitive exams. These data were obtained from the chapter on recruitment exams in the *Panoramas statistiques des personnels de l'enseignement scolaire* and the *Bilans Sociaux*, written bry the French Ministry of Education (DEPP), for the years 2010–2022.

[82] The 2013 Peillon reform modified the "masterisation" policy by moving the competitive exams from the end of the second year of the master's program to the end of the first year, which led to an increase in the number of candidates taking the exams.

volume growth is only 0.0002 points over the period 2013-2022.

In addition to novice permanent teachers, it would also be relevant to consider the share of contract teachers, who are not recruited through competitive exams and are not required to complete the same initial training as permanent teachers, which may result in lower average teaching experience (Benhenda, 2022). Under the (strong) assumption that all contract teachers do not have experience, the same quality adjustment is implemented as a robustness check in the appendix J, by aggregating the share of contract teachers with the share of novice permanent teachers. On average, between 2012 and 2021, the combined share of novice permanent teachers and contract teachers amounts to 4.2 % in primary education and 13 % in secondary education. As the share of contract teachers evolves in a similar way to that of novice teachers over the observation period (with a decline between 2012 and 2013, an increase between 2013 and 2016, and a slight decrease thereafter, see Figure 25b), the volume gaps obtained when adjusting for the combined share of contract and novice permanent teachers are very close to those obtained when adjusting only for the share of novice teachers (Figure 27).

8 Summary and comparison of results

This section summarizes and compares the results of the main quality adjustments applied (PISA scores, class size and the share of novice teachers). We consider the most plausible specifications presented above: a 10% increase in class size is associated to a -0.43% change in the volume (measured in "outcome units"), while the volume increases by 9.5% for a one-standard-deviation for PISA scores, and decreases by 0.0025% for a one-percentage-point increase in the share of novice teachers. We then discuss the consequences for productivity measures.

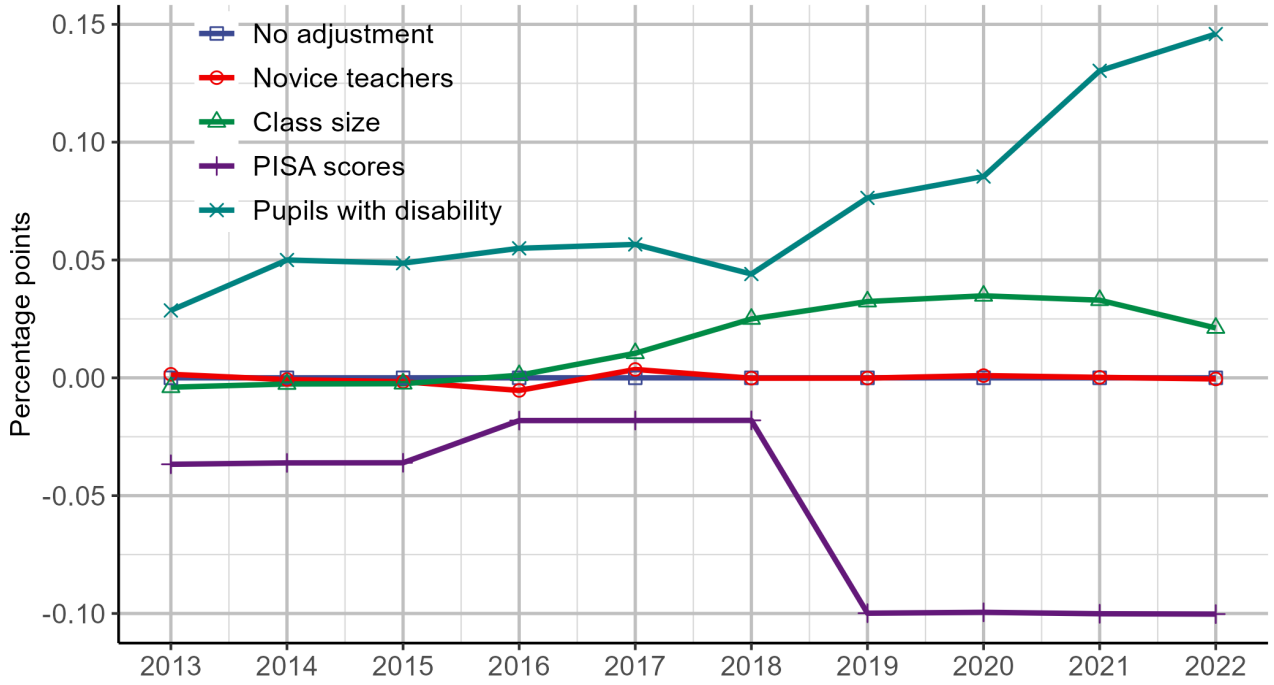
8.1 Output measures

Figure 10 shows the difference between the growth rates of the adjusted volumes and the unadjusted volume. Between 2013 and 2016, the differences in evolutions between adjusted and unadjusted volumes are modest for PISA and pupils with disability, smaller for class size and negligible for the share of novice teachers. The gap in growth rates increases from 2017 for class size, due to the reduction in class size in high-priority areas in the first two primary levels (CP and CE1) implemented that year. The adjustment for PISA scores leads to lower growth rate after 2018, as test results declined in 2022. These two adjustments display opposite dynamics relative to the unadjusted volume: the class size adjustment leads to higher volume growth, with a maximum positive gap of +0.04 percentage point in 2020. On the contrary, the PISA score adjustment results in lower growth than the unadjusted volume, with a negative gap reaching -0.1 percentage point between 2019 and 2022. The gap in growth rate is more volatile for the share of novice teachers and the maximum absolute deviation is only 0.005 percentage point.

The implicit quality adjustment using a stratification suited for pupils with disability has the strongest effect on volume, with a maximum of +0.15 percentage point in 2022. This gap remains limited but it is constantly positive, so the relative gap with the unadjusted volume expand by 0.70% over the period 2013-2022, compared to a decrease by 0.53% for the ratio of the Pisa-adjusted volume on the unadjusted volume. An adjustment can have a strong effect but still be centered around zero over a period, so it is also interesting to look at the cumulative growth differential (defined as the sum of the absolute differences between the growth rates of adjusted and unadjusted volumes). This gives similar results for PISA scores (0.56 percentage point) and disability (0.72 point) because their effect has a constant sign. The cumulative growth differential is smaller with the class size adjustment (0.17 point) and negligible for the share of novice teachers (0.01 point).

The fact that the adjustments diverge may appear counterintuitive. Accounting for the peculiarities of the education service for pupils with disability is largely independent from the other adjustments

Figure 10: Difference between the growth rates of adjusted and unadjusted volume



Note : In 2019, the difference between the growth rate of the adjusted volume by class size and the growth rate of the unadjusted volume is around 0.035 percentage point.
Coverage: First and second degree education.
Sources: DEPP, Insee, authors' computations.

that rely on students' performances and future wages. However, regarding the class size and PISA scores adjustments, one might expect them to move in the same direction, under the assumption that smaller classes improve student performance. Several factors can nevertheless explain why these quality adjustments do not converge. First, this may be due to a delayed temporal effect: the impact of the 2017 reduction in CP class sizes on the PISA scores of 15-year-old students is not yet directly observable. Second, the decline in PISA scores in 2022 may result from factors unrelated to class size. In particular, the Covid pandemic affected students' schooling and learning. Persistent socio-economic and cultural inequalities among students, as well as a less favorable disciplinary climate[83], may also explain the lower PISA scores observed in 2022 (Bernigole et al., 2023). The adjustment based on the share of novice teachers has essentially no effect, reflecting both the limited variation in the quality indicator and the calibration applied, as we explain below.

Overall, the impact of explicit quality adjustments are modest [84] due to the calibrations used to estimate the effect of quality indicators on students' future earnings. For instance, the 0.04 elasticity used for class size reduces the quality effect by a factor of 25 relative to a unitary elasticity ($f(q) = q$) calibration. These literature-based parameters strongly dampen variations in the quality adjustment variables, particularly because these adjustments are sometimes applied only to certain strata. For example, for class size, the proportion of adjusted strata represents roughly 40 % of the total value over the study period. However, those modest changes in growth rates are still important for detailed analysis, and they can aggregate into significant changes in trend if they accumulate over the years

[83]PISA produces an indicator of "disciplinary climate" based on responses to questions such as "Do students listen to teachers?" or "Is there noise or disorder in the classroom?". According to Bernigole et al. (2023), France is among the countries with one of the poorest disciplinary climates.

[84]It is worth noting that Weale (2025) finds effects of a similar order of magnitude. Applying his framework to the English Diabetes Prevention Programme, he obtains an increase in the volume of health service output of 2×10^{-2} per cent in the central accounts.

and change the interpretation of long-term evolutions, as for the implicit quality adjustment for pupils with disability that increases the total growth of education volume by one percentage point between 2006 and 2023, from -0.5% to 0.5%. This underlines the importance of having access to long time series without structural breaks to perform quality adjustment.

Table 3: Summary of the main explicit quality adjustments

	PISA Scores	**Class Size**	**Novice Teachers**
Unit of the quality indicator q	Score (points, mean = 500)	Average number of students per class	Share of novice permanent teachers (0-1 scale)
Average annual change in q (in absolute terms)	2 points	1%	0.6 point
Effect of an increase in the quality indicator	+9.5% per standard deviation increase	-0.4% when q increases by 10%	-0.0025% per percentage point increase in q
Adjusted strata	lower secondary	pre-primary and primary	all
Weights of adjusted strata	28%	40%	100%
Temporal coverage of data	2000–2022, every 3 years	2009–2024	2012–2022
Average contribution of the adjustment(*) (2013–2022)	-0.056 point	0.015 points	-0.0002 points
Maximum contribution, absolute value, of the adjustment(*) (2013–2022)	0.1 points	0.035 points	0.005 points
Recent trend of the indicator	decreasing	decreasing	decreasing
Likely future trend of the indicator	n.a.	decreasing	n.a.
Consensus on effect in the literature	✓	partial	partial
Data reliability	✓	✓	✓
Data comparability	✓	✓	✓
Data production frequency	×	✓	✓
International comparability	✓	✓	✓
Completeness vs. robustness	complete	robust	robust
Conceptual difficulties	Tests of 15-year-old students only; influence of socio-economic and cultural factors	Heterogeneity of class sizes	

Note : (*) Contribution of the adjustment to volume changes, calculated by multiplying the weight of the adjustment in the past volume by the annual growth rate of the adjusted volume
Coverage : First and second degree education.
Sources : DEPP and Insee data, authors' computations.

Table 3 summarizes the main advantages and limitations of each explicit quality adjustment for application in national accounts (cf. Appendix A.2). First, the robustness of estimates regarding the effects of each variable on students' future earnings is best documented in the literature for PISA scores. In contrast, the effects of class size and the share of novice teachers are more debated: some studies find significant effects, while others find none. The data used for these adjustments are generally reliable, cover a sufficiently long time period, are temporally comparable, and are produced within a relatively short timeframe. However, since PISA tests are conducted every three years, complementing them with results from annual assessments would allow adjustments to be made at a more reasonable frequency. The adjustment for the share of novice teachers covers all strata while the PISA scores adjustment covers only lower secondary, and class size only pre-primary and primary levels. These indicators have the advantage of being applicable in other countries, as they are largely independent of the French education system and social context, allowing for international comparisons of adjusted volumes. Finally, certain conceptual limitations may constrain the use of some adjustments. PISA scores, for example, capture student achievement only at a single point in schooling and are partly influenced by socio-economic and cultural factors. They may also reflect other dimensions of the education system, such as teaching methods or students' socio-economic background.

Finally, we can wonder whether distinct approaches to quality adjustment can be meaningfully combined, so as to better reflect the multidimensional nature of education while alleviating the limitations associated with any single measure. This could be done either by adjusting the volume using multiple quality indicators, or by combining explicit and implicit quality adjustments - namely, by introducing a quality indicator alongside a finer stratification. The underlying idea is that combining

several quality indicators could produce a quantitative measure that is both complete and robust; that is, one that captures a large part of the effects of education on students' future earnings while remaining unaffected by external factors. Here, the indicators examined have limitations regarding completeness or robustness. For instance, PISA scores are more complete, as their impact on students' future wages is substantial and well-established in the literature, but they are less robust since they are influenced by external factors such as socio-economic and cultural context. In contrast, indicators such as class size or the share of novice teachers are arguably more robust, as they are less sensitive to contextual factors, but they explain a smaller part of outcomes, given the smaller and more debated impact on students' future earnings.

As shown in section 3.3.2 and appendix A.6, if quality indicators are independent, their adjustment functions can be easily combined : the multidimensional adjustment function is simply the product of the individual adjustment functions. However, in our case, the different dimensions of education quality that we examine do not appear to be independent. While class size may be relatively independent from teacher experience, this is not the case for class size and PISA scores, nor for PISA scores and teacher experience. As a result, combining multiple quality indicators into a single adjustment becomes more complex, as it would require estimating the effect of each indicator on the others. We chose not to pursue this line of analysis in this paper. By contrast, it would be relatively straightforward to combine explicit and implicit quality adjustments. This would involve adjusting the volume using a single quality indicator while introducing a finer stratification, for example by including specific strata for students with disabilities. Combining explicit and implicit quality adjustments would approximately be equivalent to summing their effects, thus the impact would remain modest in our case.

8.2 Productivity measures

Productivity is one of the primary applications of volume measures in the National Accounts. We therefore examine how our adjustments affect standard productivity indicators. Productivity is typically defined as the ratio of output to inputs used in the production process. The most commonly used measure is labor productivity (Schreyer and Pilat, 2001), computed as the volume of value added —i.e., output net of intermediate consumption — divided by the number of full-time equivalents. By contrast, the apparent factor productivity measure introduced in Section 5 is defined as the ratio of output volume to the volume of total inputs, including intermediate consumption.

To compute productivity, we use total output rather than value added as the numerator, since data on the volume of intermediate consumption is only available for the aggregate non-market education sector and cannot be isolated for primary and secondary education. This limitation is unlikely to drastically affect our results. In particular, the share of intermediate consumption in total costs (including higher education) is relatively small and remains stable over time, as shown in Figure 12 for the overall non-market education sector.

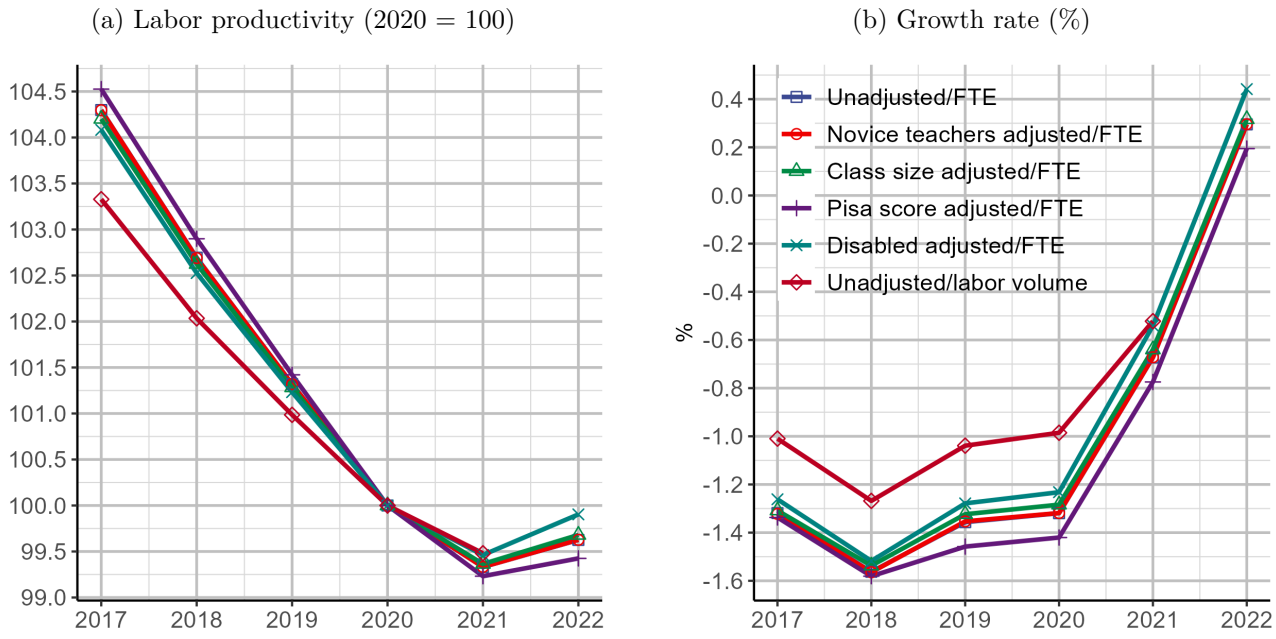
Figure 11 presents the different measures of labor productivity including the different quality adjustments. The labor productivity of first and second degree education is globally decreasing from 2017 to 2022, as the number of staff grows while the number of pupils is slightly decreasing. For small variations, the change in productivity can be approximated as the difference between the growth rate of output and that of inputs. It follows that comparing productivity measures based on alternative output indices—whether they incorporate implicit or explicit quality adjustments—is effectively equivalent to comparing the output measures themselves, as the same input growth rate is subtracted in each case. Consistent with this reasoning, we find only minor differences—below 0.15 percentage points—across labor productivity measures constructed using the various output volume indices[85]. The quality adjustments thus do not change the fact that productivity is decreasing from 2017 to 2022, which is

[85] Replacing output with value added would lead to similar gaps across productivity measures, since identical intermediate consumption would be deducted from each output measure.

also observed by Hanushek and Ettema (2017) in the US using tests scores as output. This does not imply that public policies affecting the quality of education have been useless as shown in Section A.7.

Replacing the denominator from the number of full-time equivalent employees to a Laspeyres labor index defined in Appendix C.3, has a more substantial effect on productivity growth up to 0.3 percentage points, because the composition of the education labor force has changed during the period.

Figure 11: Different measures of the labor productivity of education



Note: FTE = full-time equivalents. The productivity measures are defined with production (including intermediate consumption) as the numerator. The labor volume is a chained Laspeyres quantity index using the number of full-time equivalents and the net wages of different categories of education staff. It therefore includes compositional effects as it gives more weight to workers with higher wage.
Coverage: First and second degree education.
Source: DEPP, authors' calculation.

9 Conclusion

This study compares several quality-adjustment methods for measuring the volume of non-market education output within a conceptual framework that links the weight of each adjustment to its contribution to human capital. Our approach values education quality through an outcome variable — pupils' future earnings, which is unusual in national accounts, but this appears as the most robust procedure to avoid arbitrary choices and relies on an extensive literature. Our calibration relies on a specific set of measures capturing the fundamentally multidimensional concept of quality, as well as microeconomic estimates of the causal effects of education quality indicators on future wages. These estimates may be subject to important caveats. They are obtained in partial equilibrium and abstract from general equilibrium effects (queue effect, congestion effect).

The adjusted volumes follow similar evolutions to the unadjusted volumes, with a maximum gap in growth rates of 0.15 percentage point for the period 2013–2022. This maximum gap is reached with a finer stratification that accounts for the growing number of students with disabilities (+0.07 percentage point on average, +0.15 in 2022). The adjustment based on PISA scores results in slower growth relative to the unadjusted volume (-0.056 point per year on average, -0.10 from 2019 to 2022) whereas adjusting for class size yields slightly faster growth (+0.015 point per year on average, peaking at +0.035 in 2020). Adjusting for the share of novice teachers has a negligible effect. Overall, quality

adjustments exert a limited influence on estimated education volume. However, those modest gaps in growth rates can accumulate over the years and produce different trends: the stratification for students with disabilities increases the relative change in output over the whole period from 2006 to 2023 by approximately one percentage point, from -0.5 % to +0.5%.

These results should be interpreted carefully, and two complementary interpretations are possible — both of which we believe to be simultaneously valid, and which our decomposition framework helps to quantify.

The first interpretation is that the decline in PISA scores and the reduction in class sizes do have real effects on students' future prospects, but these effects are too small to be visible in an annual volume index. We do not claim that education quality changes are inconsequential for student welfare: the elasticities we use, even the smallest ones, imply real effects on future earnings. However, even under the most optimistic calibration — recovering France's 2018 PISA score would imply an annual gain of approximately 7 points, or 0.07 standard deviations, combined with an upper-bound elasticity of 0.3 and the cost share of the adjusted stratum (28 %) — the first-order contribution to annual volume growth amounts to approximately 0.6 percentage points. This does not mean that the effect is economically unimportant — it means that the scale of an annual volume index is not the right tool to measure the effects of an education policy. As Section A.7 in Appendix shows, a class size reduction reform welfare-improving even when its contribution to the volume index is negligible.

The second interpretation is that the current literature does not yet allow sufficiently robust and complete estimation — in the sense of our completeness and robustness criteria formalized in Section A.5 in Appendix — of the effect of all quality changes on students' future earnings. Several important dimensions remain outside the scope of existing estimates: the effects of education on non-cognitive skills are not captured by our indicators; general equilibrium effects — whether congestion and positional effects that would compress the aggregate elasticity, or positive externalities that would expand it — are not accounted for; and the time horizon of microeconomic studies is typically too short to capture the full career-long returns to education quality.

A third point, largely independent of the two above, is that even if the effects of quality on future earnings were perfectly estimated and found to be substantial, they would still translate into modest annual volume adjustments for a purely arithmetical reason: the contribution of quality to volume growth is the product of the cost share of the adjusted stratum, the elasticity, and the annual change in the quality indicator — three terms that are each individually small, making their product very small. This is a structural feature of the volume index framework, not a limitation specific to our quality indicators or elasticities. It implies that the volume index is not designed to capture the full welfare value of education policy, and should not be used as such. The public debate on the deterioration of French education is not contradicted by our results — it is simply not the question that an annual volume index is built to answer

Finally, it is worth noting that our approach relies on the choice of quality indicators and elasticities estimates taken from the literature. Future research could extend our analysis by incorporating additional quality indicator such as employment outcomes and estimate more robust elasticities of quality effects. More fundamentally, our methodology anchors the assessment of educational quality in the human capital paradigm, using future earnings as the primary metric. Nevertheless, workers may accept lower monetary compensation in exchange for other aspects of the job, which may include more pleasant working conditions or the alignment with their individual preferences. A high-quality education system might thus lead individuals to maximize such dimensions, even if this choice does not translate into higher observed incomes. As a consequence, alternative approaches might explore dimensions of educational quality beyond the scope of human capital. Developing a multidimensional framework to capture these diverse aspects of quality would represent a particularly promising avenue for further study.

References

Acemoglu, D. and Autor, D. (2011). Skills, tasks and technologies: Implications for employment and earnings. In *Handbook of labor economics*, volume 4, pages 1043–1171. Elsevier.

Acemoglu, D., Gallego, F. A., and Robinson, J. A. (2014). Institutions, human capital, and development. *Annu. Rev. Econ.*, 6(1):875–912.

Afonso, A., Schuknecht, L., and Tanzi, V. (2010). Public sector efficiency: evidence for new eu member states and emerging markets. *Applied economics*, 42(17):2147–2164.

Ahlroth, S., Björklund, A., and Forslund, A. (1997). The output of the swedish education sector. *Review of Income and Wealth*, 43(1):89–104.

Altonji, J. G. and Pierret, C. R. (2001). Employer learning and statistical discrimination. *The quarterly journal of economics*, 116(1):313–350.

Angrist, J. D., Lavy, V., Leder-Luis, J., and Shany, A. (2017). Maimonides Rule Redux. NBER Working Papers 23486, National Bureau of Economic Research, Inc.

Aninat, E. and Bauer, A. (1999). 7 addressing equity issues in policymaking: Lessons from. *Economic Policy and Equity*, page 109.

Atkinson, T. (2005). Atkinson review: Final report. measurement of government output and productivity for the national accounts. Technical report, Office for National Statistics, London. Accessed April 11, 2025.

Baradji, E. (2024). Combien d'élèves devant un enseignant pendant une heure de cours dans un établissement du second degré en 2023 ?

Baysu, G., Agirdag, O., and De Leersnyder, J. (2023). The association between perceived discriminatory climate in school and student performance in math and reading: A cross-national analysis using pisa 2018. *Journal of youth and adolescence*, 52(3):619–636.

Benhenda, A. (2022). Absence, substitutability and productivity: Evidence from teachers. *Labour Economics*, 76(C).

Benhenda, A. and Grenet, J. (2015). Combien coûte le redoublement dans l'enseignement primaire et secondaire en France ?

Bernigole, V., Fernandez, A., Loi, M., and Salles, F. (2023). Pisa 2022 : la france ne fait pas exception à la baisse généralisée des performances en culture mathématique dans l'ocde. Note d'information 23.48, DEPP.

Besley, T. and Coate, S. (1997). An economic model of representative democracy. *The Quarterly Journal of Economics*, 112(1):85–114.

Biasi, B., Deming, D. J., and Moser, P. (2021). Education and innovation. Working Paper 28544, National Bureau of Economic Research.

Bils, M. and Klenow, P. J. (2000). Does schooling cause growth? *American economic review*, 90(5):1160–1183.

Bouguen, A., Grenet, J., and Gurgand, M. (2017). La taille des classes influence-t-elle la réussite scolaire ?

Bressoux, P., Kramarz, F., and Prost, C. (2009). Teachers' training, class size and students' outcomes: Learning from administrative forecasting mistakes. *Economic Journal*, 119:540–561.

Browning, M. and Heinesen, E. (2007). Class size, teacher hours and educational attainment. *The Scandinavian Journal of Economics*, 109(2):415–438.

Card, D. (2001). Estimating the return to schooling: Progress on some persistent econometric problems. *Econometrica*, 69(5):1127–1160.

Card, D. and Krueger, A. B. (1992). Does school quality matter? returns to education and the characteristics of public schools in the united states. *Journal of political Economy*, 100(1):1–40.

Charron, N. and Rothstein, B. (2016). Does education lead to higher generalized trust? the importance of quality of government. *International Journal of Educational Development*, 50:59–73.

Chetty, R., Friedman, J. N., Hilger, N., Saez, E., Schanzenbach, D. W., and Yagan, D. (2011). How does your kindergarten classroom affect your earnings? evidence from project star. *The Quarterly journal of economics*, 126(4):1593–1660.

Clotfelter, C., Ladd, H., and Vigdor, J. (2007). Teacher credentials and student achievement in high school: A cross-subject analysis with student fixed effects. *Journal of Human Resources*, 45.

Collesi, D., Guerrucci, D., Versace, D., and Zannoni, S. (2007). The use of class size and the italian method. In *OECD Workshop on Measuring Education and Health Volume Output*, Paris, France. Paper presented at the OECD Workshop on Measuring Education and Health Volume Output.

Collesi, D., Versace, D., and Zannoni, S. (2008). Measurement of non-market service: Output and outcomes. In *30th General Conference of the International Association for Research in Income and Wealth (IARIW)*, Portoroz, Slovenia.

Connolly, M. and Haeck, C. (2018). Nonlinear Class Size Effects on Cognitive and Noncognitive Development of Young Children. Working Papers 18-01, Research Group on Human Capital, University of Quebec in Montreal's School of Management.

Cour des comptes (2024). L'inclusion scolaire des élèves en situation de handicap.

Cour des comptes (2025). L'enseignement primaire : une organisation en décalage avec les besoins de l'élève. Rapport public thématique, Cour des comptes, Paris. PDF disponible : https://www.ccomptes.fr/sites/default/files/2025-05/20250520-Enseignement-primaire.pdf.

Crahay, M. (2004). Peut-on conclure à propos des effets du redoublement? *http://lst-iiep.iiep-unesco.org/cgi-bin/wwwi32.exe/[in=epidoc1.in]/?t2000=021785/(100)*, 148.

Croguennec, Y. (2025). Prévisions d'effectifs d'élèves du premier degré.

Cutler, D. M., Ghosh, K., Messer, K. L., Raghunathan, T., Rosen, A. B., and Stewart, S. T. (2022). A satellite account for health in the united states. *American Economic Review*, 112(2):494–533.

Deming, D. (2009). Early childhood intervention and life-cycle skill development: Evidence from head start. *American Economic Journal: Applied Economics*, 1(3):111–134.

Deming, D. and Silliman, M. (2025). Skills and human capital in the labor market. In *Handbook of Labor Economics*, volume 6, pages 115–157. Elsevier.

Diewert, E. (2008). The Measurement of Nonmarket Sector Outputs and Inputs Using Cost Weights. Economics working papers diewert-08-01-18-09-21-07, Vancouver School of Economics.

Dkhissi, R., Ferrari, S. S., and Zwijnenburg, J. (2025). Comparability of economic growth estimates during covid-19: Do different methods for measuring non-market output affect international comparability? OECD Statistics Working Papers 2025/02, OECD Publishing, Paris.

Dolton, P. and Gutiérrez, O. (2011). If you pay peanuts do you get monkeys? a cross-country analysis of teacher pay and pupil performance. *Economic Policy*, 26:5 – 55.

Duflo, E. (2001). Schooling and labor market consequences of school construction in indonesia: Evidence from an unusual policy experiment. *American economic review*, 91(4):795–813.

Eccles, J. S., Wigfield, A., Midgley, C., Reuman, D., Iver, D. M., and Feldlaufer, H. (1993). Negative effects of traditional middle schools on students' motivation. *The elementary school journal*, 93(5):553–574.

Erdem, C. and Kaya, M. (2024). The relationship between school and classroom climate, and academic achievement: A meta-analysis. *School Psychology International*, 45(4):380–408.

Eurostat (2013). *Système européen des comptes – SEC 2010.* Publications Office.

Eurostat (2014). Manual on measuring research and development in esa 2010. Technical Report KS-GQ-14-004-EN-N, Publications Office of the European Union, Luxembourg.

Eurostat (2016). *Handbook on prices and volume measures in national accounts – 2016 edition.* Publications Office.

Evain, F. (2025). Taille des classes du premier degré.

Fajeau, M., Grenet, J., and Laveissière, E. (2025a). L'effet des compétences scolaires sur les salaires futurs. Focus 112, Conseil d'Analyse Économique (CAE), Paris, France.

Fajeau, M., Grenet, J., Laveissière, E., and Leonetti, O. (2025b). Efficacité des politiques éducatives : sources et hypothèses de calcul. Focus 114, Conseil d'Analyse Économique. Focus accompagnant la Note du CAE n°84 « Éducation : comment mieux orienter la dépense publique ».

Fajeau, M., Grenet, J., Laveissière, E., and Leonetti, O. (2025c). Efficacité économique de la réduction de la taille des classes. Focus 113, Conseil d'Analyse Économique (CAE), Paris, France.

Falch, T., Sandsør, A. M. J., and Strøm, B. (2017). Do smaller classes always improve students' long-run outcomes? *Oxford Bulletin of Economics and Statistics*, 79(5):654–688.

Finnan, C., Schnepel, K. C., and Anderson, L. W. (2003). Powerful learning environments: The critical link between school and classroom cultures. *Journal of Education for Students Places at Risk*, 8(4):391–418.

Fraumeni, B. M., Reinsdorf, M. B., Robinson, B. B., and Williams, M. P. (2008). Price and real output measures for the education function of government: Exploratory estimates for primary & secondary education. Working Paper 14099", National Bureau of Economic Research.

Fredriksson, P., Öckert, B., and Oosterbeek, H. (2012). Long-term effects of class size *. *The Quarterly Journal of Economics*, 128(1):249–285.

Gemmell, N., Nolan, P., and Scobie, G. (2017). Public sector productivity: Quality adjusting sector-level data on New Zealand schools. Working Papers 2017/02, New Zealand Productivity Commission.

Ghaith, G. (2003). The relationship between forms of instruction, achievement and perceptions of classroom climate. *Educational research*, 45(1):83–93.

Goldberg, P. K. and Maggi, G. (1999). Protection for sale: An empirical investigation. *American Economic Review*, 89(5):1135–1155.

Goodenow, C. and Grady, K. E. (1993). The relationship of school belonging and friends' values to academic motivation among urban adolescent students. *The journal of experimental education*, 62(1):60–71.

Grossman, G. M. and Helpman, E. (1994). Protection for sale. *The American Economic Review*, 84(4):833–850.

Gu, W. and Wong, A. (2012). Measuring the economic output of the education sector in the national accounts. Technical report, Statistics Canada, Ottawa.

Hanushek, E. and Ettema, E. (2017). Defining productivity in education: Issues and illustrations. *The American Economist*, 62:056943451668820.

Hanushek, E., Rivkin, S., and Kain, J. (2005). Teachers, schools, and academic achievement. *Econometrica*, 73:417–458.

Hanushek, E. A. (2013). Economic growth in developing countries: The role of human capital. *Economics of education review*, 37:204–212.

Hanushek, E. A., Machin, S., and Woessmann, L., editors (2023). *Chapter 6 - Educational inequality*, volume 6 of *Handbook of the Economics of Education*. Elsevier.

Hanushek, E. A. and Woessmann, L. (2006). Does educational tracking affect performance and inequality? differences-in-differences evidence across countries. *The Economic Journal*, 116(510):C63–C76.

Hanushek, E. A. and Woessmann, L. (2012). Schooling, educational achievement, and the latin american growth puzzle. *Journal of Development Economics*, 99(2):497–512.

Hanushek, E. A. and Wößmann, L. (2007). The role of education quality in economic growth.

Haveman, R. H. and Wolfe, B. L. (1984). Schooling and economic well-being: The role of nonmarket effects. *The Journal of Human Resources*, 19(3):377–407.

Heckman, J. J., Lochner, L. J., and Todd, P. E. (2006). Earnings functions, rates of return and treatment effects: The mincer equation and beyond. *Handbook of the Economics of Education*, 1:307–458.

Hendren, N. and Sprung-Keyser, B. (2020). A Unified Welfare Analysis of Government Policies. *Quarterly Journal of Economics*, 135(3):1209–1318. Publisher: Oxford University Press / USA.

Heys, R. and Taylor, C. (2025). Inclusive income accounts: Practical measures to go beyond gdp. *Review of Income and Wealth*, 71(3):e70025.

Hoxby, C. (2000). The effects of class size on student achievement: New evidence from population variation. *The Quarterly Journal of Economics*, 115:1239–1285.

Insee (2025). Les comptes de la nation en 2024.

Jeljoul, M., Landreau-Mascaro, A., Dalous, J.-P., and Lemerle, S. (2016). Le compte de l'éducation principes, méthodes et résultats pour les années 2006 à 2014. Technical report, DEPP.

Jorgenson, D. and Fraumeni, B. M. (1989). The Accumulation of Human and Nonhuman Capital, 1948-84. In *The Measurement of Saving, Investment, and Wealth*, NBER Chapters, pages 227–286. National Bureau of Economic Research, Inc.

Katz, L. F. and Murphy, K. M. (1992). Changes in relative wages, 1963–1987: supply and demand factors. *The quarterly journal of economics*, 107(1):35–78.

Kline, P. and Walters, C. R. (2016). Evaluating public programs with close substitutes: The case of head start. *The Quarterly Journal of Economics*, 131(4):1795–1848.

Krueger, A. B. (1999). Experimental estimates of education production functions*. *The Quarterly Journal of Economics*, 114(2):497–532.

Lazear, E. P. (2003). Teacher incentives. *Swedish economic policy review*, 10(2):179–214.

Leuven, E. and Løkken, S. A. (2020). Long-term impacts of class size in compulsory school. *Journal of Human Resources*, 55(1):309–348.

Leuven, E., Oosterbeek, H., and Rønning, M. (2008). Quasi-experimental estimates of the effect of class size on achievement in norway. *Scandinavian Journal of Economics*, 110(4):663–693.

List, J. A. and Uchida, H. (2024). Here today, gone tomorrow? toward an understanding of fade-out in early childhood education programs. Technical report, National Bureau of Economic Research.

Lucas Jr, R. E. (1988). On the mechanics of economic development. *Journal of monetary economics*, 22(1):3–42.

Mayshar, J. (1990). On measures of excess burden and their application. *Journal of Public Economics*, 43(3):263–289.

Miconnet, N. (2025). Prévisions des effectifs du second degré pour les années 2025 à 2029.

Mitchell, J., Lewis, J., Zwijnenburg, J., Dkhissi, R., and Prendergast, T. (2022). International comparisons of the measurement of non-market output during the covid-19 pandemic.

Montenegro, C. E. and Patrinos, H. A. (2023). A data set of comparable estimates of the private rate of return to schooling in the world, 1970–2014. *International Journal of Manpower*, 44(6):1248–1268.

Mulligan, C. B. (1999). Galton versus the human capital approach to inheritance. *Journal of political Economy*, 107(S6):S184–S224.

Murnane, R. J., Willett, J. B., Duhaldeborde, Y., and Tyler, J. H. (2000). How important are the cognitive skills of teenagers in predicting subsequent earnings? *Journal of Policy Analysis and Management*, 19(4):547–568.

Murray, R. (2010). Measuring output from the public sector: a critical examination of the atkinson review. *Review of Income and Wealth*, 56(2):413–423.

Nye, B. and Hedges, L. (2004). How large are teacher effects? *Educational Evaluation and Policy Analysis*, 26:237–257.

OECD (2022). PISA 2022 Technical Report. Technical report, OECD Publishing, Paris. Revised version, May 2025.

OECD (2025). *Education at a Glance 2025: OECD Indicators*. OECD Publishing, Paris.

Okubo, T. (2022). Theoretical considerations on scaling methodology in pisa. *OECD Education Working Papers*, (282):0_1–31.

O'Mahony, M. and Timmer, M. P. (2009). Output, input and productivity measures at the industry level: The eu klems database. *The Economic Journal*, 119(538):F374–F403.

ONS (2025). National statistician's independent review of the measurement of public services productivity. Technical report, UK Statistics Authority, London. Accessed April 11, 2025.

Piketty, T. and Valdenaire, M. (2006). *L'impact de la taille des classes sur la réussite scolaire dans les écoles, collèges et lycées français - Estimations à partir du panel primaire 1997 et du panel secondaire 1995*. Ministère de l'éducation nationale.

Psacharopoulos, G. and Patrinos, H. A. (2018). Returns to investment in education: a decennial review of the global literature. *Education Economics*, 26(5):445–458.

Schreyer, P. (2010). Towards measuring the volume output of education and health services: A handbook. OECD Statistics Working Papers 2010/02, OECD Publishing, Paris.

Schreyer, P. (2012). Output, outcome, and quality adjustment in measuring health and education services. *Review of Income and Wealth*, 58(2):257–278.

Schreyer, P. and Pilat, D. (2001). Measuring productivity. *OECD Economic studies*, 33(2):127–170.

Sen, A. (1997). Editorial: Human capital and human capability. *World Development*, 25(12):1959–1961.

Shindler, J., Jones, A., Williams, A. D., Taylor, C., and Cardenas, H. (2016). The school climate-student achievement connection: If we want achievement gains, we need to begin by improving the climate. *Journal of School Administration Research and Development*, 1(1):9–16.

Shleifer, A. and Vishny, R. W. (1993). Corruption. *The quarterly journal of economics*, 108(3):599–617.

Spence, M. (1973). Job market signaling. *The Quarterly Journal of Economics*, 87(3):355–374.

Stefanou, A. (2025). La part du temps d'enseignement non assuré dans les établissements publics du second degré en 2023-2024.

Stiglitz, J. E., Sen, A., and Fitoussi, J.-P. (2009). Report by the commission on the measurement of economic performance and social progress. Technical report, Commission on the Measurement of Economic Performance and Social Progress, Paris.

Tenne, Y., Loeffel, L., Moreau-Fauvarque, C., Canvel, A., Fauchier-Magnan, E., Furnon-Petrescu, H., and Gady-Cherrier, C. (2023). Acte II de l'école inclusive. Technical Report IGÉSR n° 22-23 107B – IGAS n° 2022-114R, Inspection générale des affaires sociales (IGAS) et Inspection générale de l'éducation, du sport et de la recherche (IGÉSR).

Thomas, F. (2025). Résultats définitifs de la session 2024 du baccalauréat : un taux de réussite en hausse dans chacune des voies. *Note d'Information*, (25-05).

Thurow, L. C. (1975). Job competition: the job distribution. In *Generating Inequality*, pages 98–128. Springer.

Tomaszewski, W., Xiang, N., and Huang, Y. (2024). School climate, student engagement and academic achievement across school sectors in australia. *The Australian Educational Researcher*, 51(2):667–695.

Triplett, J. (2006). Handbook on hedonic indexes and quality adjustments in price indexes: Special application to information technology products.

UNECE (2017). *Guide on Measuring Human Capital*. United Nations Economic Commission for Europe.

Voight, A. and Hanson, T. (2017). How are middle school climate and academic performance related across schools and over time? rel 2017-212. *Regional Educational Laboratory West*.

Weale, M. (2025). New medical treatments, quality improvements and public health measures in the national accounts: the effects of the diabetes prevention programme. *Review of Income and Wealth*, 71(2):e70014.

Yu, Z. and Jiang, X. (2022). How school climate affects student performance: A fuzzy set qualitative comparative analysis based on the perspective of education system. *Mathematical Problems in Engineering*, 2022(1):7975433.

Appendices

A Conceptual framework

A.1 International conventions

	ESA 2010 (Handbook 2016)	ESA 1995 (Handbook 2001)	SNA
Input	not recommended (C)	not recommended (C)	possible
Output (unadjusted)	recommended (A)	possible (B)	recommended
Output (adjusted)	forbidden	recommended (A)	recommended*
Outcome	forbidden	forbidden	forbidden

Table 4: International conventions regarding non-market education volume measure

Note: The different edition of the European System of Account and the corresponding Handbooks on prices and volume measures in national accounts attribute grades A,B or C to the different method (A is the best and C is the worst)
*The SNA 2008 is not clearly advocating for quality adjustment but it states that they should be used when it is necessary: "Secondly, quantity indicators should be adjusted for quality change. For example, services should be sufficiently differentiated with the aim of arriving at categories that can be regarded as homogeneous [...]. In addition, the quantity indicator of each category can be augmented by an explicit quality adjustment factor", Paragraph 15.122

A.2 National accounting constraints on data

Following article 12 of the regulation on European Statistics[86] and the guide from the ONS[87], we identify a set of key properties that time series data used in national accounts should satisfy:

1. **Reliability**: indicators should be unbiased, have small standard error, and not be based on strong assumptions.
2. **Consistency over time**: indicators should be stable over time, with limited revisions and methodological breaks.
3. **Adequate frequency**: data should be available at least annually; lower frequencies may be acceptable if missing observations can be robustly imputed.
4. **Long-term availability**: indicators should be available over a sufficiently long historical period and remain available in the future.
5. **Comprehensiveness**: indicators should cover all relevant strata of the activity considered.
6. **Timeliness**: data for year t should be released within one to two years ($t+1$ or $t+2$).
7. **International comparability**: cross-country comparability is crucial for national accounts, it would likely be a key criteria for the European System of Accounts (ESA) to allow quality adjustments in the future

[86] Regulation No 223/2009
[87] A guide to quality adjustment in public service productivity measures, Methodological and statistical properties

A.3 Decomposition of production price into input price and productivity

We start with the decomposition of the output price index from Schreyer (2012), still assuming that inputs quantities are chosen to minimize costs:

$$P^t_{output} = \frac{C^t(y^t, w^t)}{C^{t-1}(y^t, w^{t-1})} = \overbrace{\frac{C^t(y^t, w^t)}{C^t(y^t, w^{t-1})}}^{\text{only } w \text{ changes}} \overbrace{\frac{C^t(y^t, w^{t-1})}{C^{t-1}(y^t, w^{t-1})}}^{\text{only technology changes}} \tag{13}$$

We cannot compute this decomposition directly because the term $C^t(y^t, w^{t-1})$, the cost of producing quantities y^t with the technology available at t and the input prices of $t-1$ is a counterfactual which cannot be observed. We assume that the production functions has constant return to scale $y^t_i = A^t F_i(X^t_{i,1}, ...X^t_{i,M})$, the technology term is equal to the growth rate of the inverse total factor productivity: $\frac{C^t(y^t,w^{t-1})}{C^{t-1}(y^t,w^{t-1})} = \frac{A^{t-1}}{A^t}$. We have:

$$\frac{C^t(y^t, w^t)}{C^t(y^t, w^{t-1})} = \frac{\sum_i X^t_i(y^t_i, w^t) \cdot w^t}{\sum_i X^t_i(y^t_i, w^{t-1}) \cdot w^{t-1}} = \overbrace{\frac{\sum_i X^t_i(y^t_i, w^t) \cdot w^t}{\sum_i X^t_i(y^t_i, w^t) \cdot w^{t-1}}}^{\text{price index of input } P^t_{input}} \overbrace{\frac{\sum_i X^t_i(y^t_i, w^t) \cdot w^{t-1}}{\sum_i X^t_i(y^t_i, w^{t-1}) \cdot w^{t-1}}}^{\text{substitution of inputs } sub^t} = P^t_{input} \times sub^t$$

In general, $\frac{C^t(y^t,w^t)}{C^t(y^t,w^{t-1})}$ is not equal to the Paasche index of input price P^t_{input} because the quantity of each input evolve when their price change, even if the technology is unchanged, so $X^t_i(y^t_i, w^{t-1}) \neq X^t_i(y^t_i, w^t)$. The term sub^t represents the increase in cost due to this substitution between inputs with technology held constant. Since $X^t_i(y^t_i, w^{t-1})$ is the optimal choice of inputs when input prices are w^{t-1}, we have $X^t_i(y^t_i, w^{t-1}) \cdot w^{t-1} \leq X^t_i(y^t_i, w^t) \cdot w^{t-1}$ thus $sub^t \geq 1$. The production function F_i determines how inputs can substitute and therefore affect the sub^t term.

A simple example is when the production function is a Leontief defined by the input-output table IO^t_i. In this case, the quantity of each input $X^t_i(y_i, w) = y^t_i IO^t_i$ is independent of input price so there is no substitution effect:

$$sub^t = \frac{\sum_i X^t_i(y^t_i, w^t) \cdot w^{t-1}}{\sum_i X^t_i(y^t_i, w^{t-1}) \cdot w^{t-1}} = \frac{\sum_i y^t_i IO^t_i \cdot w^t}{\sum_i y^t_i IO^t_i \cdot w^t} = 1$$

An example where substitution occurs is the a Cobb-Douglas $y_i = A^t (X^t_{i1})^\alpha (X^t_{i2})^{1-\alpha}$ if there are only two inputs. Here, the elasticity of substitution between inputs is equal to one, so if w_1/w_2 marginally increases by 1% while technology A^t is unchanged, X_{i1}/X_{i2} decreases by 1%. To illustrate the effect of sub^t, let us present an example of a large change in relative price that leaves the Paasche price index unchanged. We consider a single product i so we drop the index. We assume that the production function is a Kobb-Douglas with $\alpha = 1/2$, a constant total factor productivity $A^{t-1} = A^t$ and a constant production $y^t = y^{t-1}$, thus the quantity of each input depends only on their prices. We assume those evolve in opposite directions $w^{t-1}_1 = w^{t-1}_2 = 1, w^t_1 = 2$ and $w^2_t = 2/3$. Solving the maximization problem naturally yields equal amounts of inputs in the first period ($X^{t-1}_1 = X^{t-1}_2 = 1$), a decrease in input 1 ($X^t_1 = 1/\sqrt{3} < 1$) and an increase in input 2 ($X^t_2 = \sqrt{3} > 1$). Those opposite movements compensate in the Paasche price index: $P^t_{input} = \frac{X^t_1 w^t_1 + X^t_2 w^t_2}{X^t_1 w^{t-1}_1 + X^t_2 w^{t-1}_2} = \frac{1/\sqrt{3}\times 2 + \sqrt{3}\times 2/3}{1/\sqrt{3}\times 1 + \sqrt{3}\times 1} = 1$. Nevertheless, $sub^t = \frac{X^t_1 w^{t-1}_1 + X^t_2 w^{t-1}_2}{X^{t-1}_1 w^{t-1}_1 + X^{t-1}_2 w^{t-1}_2} = \frac{1/\sqrt{3}\times 1 + \sqrt{3}\times 1}{1\times 1 + 1\times 1} = \frac{2}{\sqrt{3}} > 1$. In this example, the price index of input is constant, but the change in relative price involves a substitution between inputs that raises the required amount of inputs ($Q_{input} = \frac{2}{\sqrt{3}} > 1$).

The substitution is not captured by P^t_{input} as it relies solely on quantities at time t, and it is unobserved because both technology and input prices evolve in each period so we cannot measure the counterfactual $X^t(y^t_i, w^{t-1})$. Consequently, we cannot compute $\frac{A^{t-1}}{A^t}$ and sub^t, but we can deduce their

product as a residual $\frac{P^t_{output}}{P^t_{input}} = \frac{A^{t-1}}{A^t} \times sub^t$. Since $Value^t = Q^t_{input} \times P^t_{input} = Q^t_{output} \times P^t_{output}$ this is also equal to $\frac{Q^t_{output}}{Q^t_{input}}$.

The ratio $\frac{Q^t_{output}}{Q^t_{input}}$ is a natural notion of productivity that is easy to compute with national accounts. We call it the *apparent factor productivity* $\frac{AFP^t}{AFP^{t-1}} = \frac{Q^t_{output}}{Q^t_{input}}$, and we can link it to *total factor productivity*:

$$\frac{AFP^t}{AFP^{t-1}} = \frac{Q^t_{output}}{Q^t_{input}} = \frac{P^t_{input}}{P^t_{output}} = \frac{A^t}{A^{t-1}} \times \frac{1}{sub^t}$$

This apparent factor productivity can be seen as a productivity in a broad sense, as substitution across inputs is part of the production process. The growth rate of the apparent factor productivity is a lower bound for the evolution of TFP as $AFP^t \leq \frac{A^t}{A^{t-1}}$. Assuming like Diewert (2008) that the production function is a Leontief, there is no substitution effect so we can exactly measure the evolution of total factor productivity with $\frac{AFP^t}{AFP^{t-1}} = \frac{Q^t_{output}}{Q^t_{input}} = \frac{A^t}{A^{t-1}}$. The main input of the education service is labor, which seems weakly substitutable in the production process, so the Leontief assumption is reasonable.

A.4 Earnings-based quality adjustment and the double-counting concern

In our framework, a quality improvement raises the volume of education output today — through the substitution rate between quality and quantity in the index — and also raises overall production later, as better-trained workers enter the labour market. We argue that this does not constitute double counting from a national accounts perspective, since the two records at issue concern different transactions in different periods. It is analogous to a quality gain in a capital good: a more powerful computer is recorded as a volume increase in current capital formation, and in turn raises the volume of production in later years by making workers more productive. Similarly, for a market education service e.g. in higher education, a conventional quality adjustment such as a hedonic regression would capture the fact that households are willing to pay more for higher quality education if it results in higher future wages. In all cases, there are two distinct production processes which are linked, but each produces a given quality-adjusted amount of product that should be recorded in the national accounts.

The main difference between human and non-human capital is that non-human capital is explicitly measured in national accounts so capital formation and capital services are clearly identified, whereas there is no human capital in the central framework, so education is considered as final consumption. However, the fact that it is not measured does not imply that the willingness to pay for education is not determined by future gains from this service, and the volume measure should capture this willingness to pay as described in Section 3.1.

In terms of productivity analysis, an increase in the quality of capital leads to higher labor productivity due to capital deepening, which can be separated from the variation in total factor productivity (TFP) by controlling for the volume of capital. In the same way, the contribution of human capital to labor productivity growth can be distinguished by controlling for educational composition and quality of education, otherwise this contribution is absorbed in the TFP. A rise in the quality of education can have either a positive or negative effect on the productivity of the education service itself, depending on the inputs necessary for this quality improvement. It has a positive effect on future labor productivity, but this effect should theoretically disappear when controlling for the variation in human capital.

A second difference with the hedonic case concerns the timing of the signal. Every other valuation in the central framework rests on preferences observed contemporaneously, whereas here the effect of quality on volume is justified by an event occurring often two decades later[88]. Resolving this

[88] Apprentices likely use professional skills learned at school within months, so the temporal gap is not necessarily large.

tension would require incorporating a human capital asset on the balance sheet, with a corresponding consumption of fixed capital.

The nominal value of education output and each of its components is identical with and without the adjustment; only the split of a fixed nominal envelope between volume and price is affected. Therefore, in terms of value, only the value of future production is affected by the quality change in education, so there is clearly no double counting. This double-counting issue would be a point to consider in building a complete framework where human capital is measured using future wages, as in **?**.

Our explicit quality adjustment based on future earnings is therefore internally coherent as a volume measure, but it rests on a prospective preference signal whose balance-sheet counterpart cannot exist within the central framework. This is why it belongs more naturally in supplementary tables or a satellite account, whereas implicit adjustment through stratification — where the cost differential is contemporaneous and observable — raises none of these difficulties.

A.5 Completeness and robustness

The completeness of a quality indicator q can be expressed in the notations of our conceptual framework as the share of the variation outcome that are due to change in q by $completeness = \frac{\partial H}{\partial q}\frac{\Delta q}{\Delta H}$. A quality indicator is more complete if it explains a larger part of the variation of outcome.

To introduce robustness, we need to express a quality indicator as not intrinsically linked to the public service, but as a proxy or a consequence of the characteristics of the service itself. We write $q = g(X, \Omega)$ where q is the measured quality indicator, X is the underlying characteristics of non-market production such that $\frac{\partial q}{\partial X} = 1$ thus $dq = dX + \frac{\partial q}{\partial \Omega}d\Omega$. The variable X represents the vector of all the underlying characteristics of the service that affect quality even if they are not measured, normalized such that $\frac{\partial q}{\partial X} = 1$. For example, if q is the grade, X is the variables of the education system that explain the variation in scores. Therefore, X would be a better quality indicator as it is independent of external factors Ω, but X cannot be measured directly. By definition, the effect of X on outcome is the same as that of q ($\frac{\partial H}{\partial X} = \frac{\partial H}{\partial q}\frac{\partial q}{\partial X} = \frac{\partial H}{\partial q}$). However, we only know the variations of q, $dq = dX + \frac{\partial q}{\partial \Omega}d\Omega$, so we need that $\frac{\partial q}{\partial \Omega} \approx 0$ so that $dq \approx dX$. We thus define robustness as the fact that the external factors causes only a small part of the variations of the quality indicator: $\frac{\partial q}{\partial \Omega}d\Omega \ll dq$[89].

A.6 Combining several adjustments

To perform a quality adjustment adjust using multiple quality indicators, the definition of the adjustment function f is more complex because it needs to reflects how the dimensions of quality interact in their effect of outcome. Intuitively, if two variables q_1 and q_2 have independent effect on outcome (for example class size and teacher's experience, to a certain extent), their effect can simply be combined. Conversely, if q_1 affects H partly through q_2 (for example class size and grades), then they cannot be brought together without additional treatment.

To formalize this, we denote[90] $q = (q_1, ..., q_m)$ and $\tilde{H} = \tilde{H}(\tilde{\Omega}, q_1, ..., q_m)$ the outcome. Let $1 \leq j \leq m$, and assume q_j is for example class size. We use tilde for H and Ω because they are not exactly equal to the outcome function and external factor for a single quality indicator q_j defined in section 3.3.2. We denote $H_j(q_j, \Omega_j)$ the outcome function for a *single* quality indicator q_j. We always have $H_j(q_j, \Omega_j) = \tilde{H}(\tilde{\Omega}, q_1, ..., q_n)$, but Ω_j can contains some other dimensions of quality q_l for $l \neq j$ (class size q_j is in Ω_l if q_l is teacher's experience), whereas $\tilde{\Omega}$ represents all the variable that affect outcome besides $q_1, ..., q_m$. The function $\tilde{H}$ is also different because q_j can affect the outcome through another indicator q_k (class size q_j affect future outcome through grades q_k), therefore the derivative $\frac{\partial \tilde{H}(\tilde{\Omega}, q_1, ..., q_m)}{\partial q_j}$ is a

[89] This is equivalent to $\frac{\partial q}{\partial \Omega}d\Omega \ll dX$, but dX in unobserved so we express the condition in term of dq only

[90] Here the subscript is not the stratum but the dimensions of quality, we drop the subscript index i of stratum for readability

priori not equal to $\frac{\partial H_j(\Omega,q_j)}{\partial q_j}$, because in $\tilde{H}$, the partial derivative holds every other quality dimensions constant. For example, the effect of class size on future wage holding grades constant $\frac{\partial \tilde{H}}{\partial q_j}$ is likely zero, even though the effect of class size of wage $\frac{\partial H_j}{\partial q_j}$ is not zero. Since $H_j(q_j,\Omega) = \tilde{H}(\tilde{\Omega}, q_1, ..., q_m)$, we obtain:

$$\frac{\partial H_j}{\partial q_j} = \frac{\partial \tilde{H}}{\partial q_j} + \sum_{k \neq j} \frac{\partial \tilde{H}}{\partial q_k} \frac{\partial q_k}{\partial q_j}$$

This formula reflects the fact that a quality variable affect outcome directly but also indirectly through other dimensions of quality. The literature usually provide estimates for $\frac{\partial H_j}{\partial q_j}$, but not for $\frac{\partial \tilde{H}}{\partial q_j}$. In addition, the $\frac{\partial q_k}{\partial q_j}$ are also hard to pin down. If all quality dimensions are independent, meaning $\frac{\partial q_k}{\partial q_j} = 0$ for $k \neq j$, then $\frac{\partial H_j}{\partial q_j} = \frac{\partial \tilde{H}}{\partial q_j}$ and

$$d\tilde{H} = \sum_j \frac{\partial \tilde{H}(\tilde{\Omega}, q_1, ..., q_m)}{\partial q_j} dq_j = \sum_j \frac{\partial H_j(\Omega, q_j)}{\partial q_j} dq_j$$

The quality adjustment functions f_j for a single quality indicator q_j are such that $\frac{f_j'(q_j)}{f_j(q_j)} = \frac{\partial H_j / H_j}{\partial q_j}$. Hence, if all quality dimensions are independent, we can simply define the multidimensional adjustment function as the product of the individual adjustment functions $f(q_1, ..., q_m) = \prod_{j=1}^m f_j(q_j)$, so that the marginal variation of f is still equal to the marginal variation of the outcome:

$$\frac{df(q)}{f(q)} = \sum_j \frac{f_j'(q_j)}{f_j(q_j)} dq_j = \sum_j \frac{\partial H_j}{\partial q_j} \frac{dq_j}{H_j} = \frac{1}{\tilde{H}} \sum_j \frac{\partial H_j}{\partial q_j} dq_j = \frac{d\tilde{H}}{\tilde{H}}$$

A.7 Volume measures and public policy

One natural consequence of our work for policymakers would be to use volume of value added as an evaluation tool and a guide for resource allocationn. However, while productivity (output per unit of input) is a relevant metric for tracking organizational efficiency, it is not the appropriate tool for evaluating whether a public policy is worthwhile. To guide resource allocation, one must instead rely on the Marginal Value of Public Funds (MVPF), which directly measures whether a policy creates more welfare than it costs. The discrepancy between these metrics can be substantial: a policy may simultaneously reduce average productivity and generate welfare gains exceeding its net cost, meaning a reliance on productivity alone would lead to rejecting good policies. To illustrate these discrepancies, we compare the quality-adjusted volume measures of output and productivity with the MVPF for a policy affecting the same quality indicator.

The MVPF (Mayshar, 1990; Hendren and Sprung-Keyser, 2020) of a policy is defined as the ratio of the change in welfare in monetary terms (the willingness to pay) caused by the policy to its net cost for the government:

$$MVPF = \frac{\frac{dW}{dq}}{\frac{dC}{dq} - \frac{dE}{dq}}$$

where W is the welfare of the pupil proxied by the discounted sum of future incomes, C is the gross cost of implementing the policy for the government, and E is the fiscal externality such as higher tax revenues or lower social spending in the future if the policy increases future income. It can be applied both to public expenses and taxes, and it includes the behavioral responses of the agents such as the distortionary effects of taxes or the future fiscal gains due to higher income. Welfare maximization for the government implies prioritizing policies with the highest MVPF: a policy with MVPF higher than 1 is generally welfare-increasing as a lump-sum tax has an MVPF of one. Fajeau et al. (2025b)

computes the MVPF of different education policies for France on a per-student basis with this formula.

As outlined in Section 3.3.3, the marginal variation of the quality-adjusted chained-linked volume of education for a single stratum is

$$\frac{dQ_{adj}^{chained}}{Q_{adj}^{chained}} = \frac{dy}{y} + \frac{df(q)}{f(q)} = \frac{dy}{y} + \frac{dW}{W}$$

Therefore, the variation in productivity defined broadly as $\Pi = Q_{adj}^{chained}/C$ is:

$$\frac{d\Pi}{\Pi} = \frac{dy}{y} + \frac{dW}{W} - \frac{dC}{C}$$

Both the MVPF and the change in productivity reflect the effect of quality on welfare and the cost of a quality improvement. Nevertheless, there are important differences between those two indicators.

A first element is that volume and productivity are strongly affected by the number of pupils y so the consequence of a quality change can be negligible if it is small compared to the impact of demography, whereas the MVPF only focuses on the consequences of a quality change through a specific policy. For example, the annual variation in the number of pupils dy/y is around ± 0.5 % while the effect of quality on wage $\frac{\partial W}{\partial q}dq$ is in the range of ± 0.05 % or smaller for class size and PISA scores. The MVPF is therefore closer to the *contribution of quality* to productivity change (i.e. assuming $dy = 0$):

$$\left.\frac{d\Pi}{\Pi}\right|_y = \frac{\partial W}{\partial q}dq - \frac{dC}{C}$$

There are still important differences between this contribution and the MVPF. First, the MVPF includes the fiscal externality $dE = \tau dW$ due to the higher future wages with τ the marginal tax rate, whereas the productivity only relates to the inputs used for production. This is pivotal since this fiscal externality is comparable or higher than the gross cost for several education policy according to Fajeau et al. (2025b): class size reduction in first degree is self-financed because $dC = 2856 < dE = 2934$ so $MVPF = +\infty$. Productivity measures how much is produced per unit of inputs, for a public or private sector activity, so it should not take into account those externalities that are paramount for policy evaluation.

Even if we omit the fiscal externality term dE in the MVPF, it provides a different insight than the change in production or productivity because it does not say something about the *relative* change of output and input, but only the *absolute* marginal variations: an increase in q can have a small relative effect on welfare dW/W (and thereby on output volume) and a high relative effect on cost dC/C, but the increase in welfare dW can still be higher than the marginal cost dC if W/C is high. Fajeau et al. (2025c) evaluates the discounted value of future wages of a pupil at 441,356 euros, so splitting a class of 22 in two increases the wages of pupils by 4,539 euros, which is only a 0.01 percent in output volume. However, it costs only 2,856 euros per pupil, the wage of the additional teachers divided by 22. In most cases, the marginal gains of quality are decreasing $\frac{\partial^2 W}{\partial q^2} < 0$ as for a production factor in a standard model, so the average productivity is decreasing with q ($d\Pi < 0$). Nonetheless, if the additional welfare is higher than the marginal cost, a government with a strictly intertemporal objective function (neglecting annual budgetary constraints) should keep investing in quality just like a private company increases spending until marginal productivity is equal to price.

In summary, a policy that has negligible effects on output volume and decreases the average productivity of the public service, can still lead to an increase in welfare higher than its net cost and even be Pareto-improving if the indirect fiscal effects compensate the cost. This is why productivity, while useful for operational management, cannot serve as the guide for policy evaluation—the MVPF is.

B Summary table of the methods tested

Volume measure	Data/Disaggregation	Period	Sources	Coverage	Section
Input method					
Input-price-based volume	Production costs (wages, intermediate consumption, capital consumption) deflated using specific price indices, detail by level unavailable	2013–2022	Insee (National Accounts)	All levels including higher education.	4.2
Market-price input volume	CPI for market-based education services (COICOP), details by level unavailable	2013–2022	Insee (National Accounts)	All levels including higher education	4.2
Output approach without quality adjustment					
Number of pupils	Number of pupils and projected number of pupils by level, expense of general government for teaching by level	2010–2028	DEPP	Pre-primary to upper secondary. Higher education excluded⋆.	5.1
Number of pupil-hours	Teaching hours per pupil (E/S × H/E for secondary education) by level of secondary education (no data for special secondary)	2011–2023	DEPP	Pre-primary to upper secondary. Higher education excluded.	5.2
Output approach with implicit quality adjustment					
Stratification adapted to pupils with disability	Number of pupils with disabilities by level† and total expenditure for those pupils	2006–2023	DEPP, Budget report (*RAP*, Program 230)	Pre-primary to upper secondary. Higher education excluded.	6
Output approach with explicit quality adjustment					
PISA-scores adjustment	Average French PISA scores (reading, mathematics, science)	2010–2022	OECD	Pre-primary to upper secondary. Higher education excluded. Adjustment for lower secondary only.	7.1
Class-size adjustment	Average class size (E/S indicator for secondary education)	2010–2023	DEPP	Pre-primary to upper secondary. Higher education excluded. Adjustment for pre-primary and primary only.	7.2
Novice-teacher adjustment	Share of zero-experience teachers estimated from recruitment exam data	2012–2022	DEPP	Pre-primary to upper secondary. Higher education excluded. Adjustment for all levels.	7.3

⋆ The output volume used in comparison to the input volume in Section 4 includes higher education (post-secondary non-tertiary, short tertiary education, long tertiary education)
† The number of pupils with disabilities in primary education is only available by age, so pre-primary and primary levels are inferred from age groups.

C Input method

C.1 Data

The input method requires a price index for each production cost: compensation of employees (D1), which includes both gross wages and salaries and employer social contributions; intermediate consumption (P2), i.e., goods and services used as inputs during production, excluding fixed assets whose consumption is recorded as consumption of fixed capital (P51C). We employ the civil servant gross wage index [91] as the price index for compensation of employees. Note that National Accounts can make more precise distinctions by decomposing bonuses into a price component and a volume component[92]. For intermediate consumption, and following the method used by the French National Accounts Department, we deflate intermediate consumption for each product by a corresponding price index. For consumption of fixed capital, we use a price index specific for each type of asset.

To compute directly the laspeyres quantity index of labor, we use data on the number of employees and average gross wage from the statistical office of the French Ministry of Education (DEPP) and the social balance sheets database [93].

With respect to non-teaching staff, our analysis is restricted to public-sector employees, without distinguishing education levels. Due to two methodological breaks occurring in 2016 and 2022, the study period is limited to the years 2016–2021 [94].

[91] This index aims at measuring the changes in the gross wage of civil servants, for each category (A, B, and C).

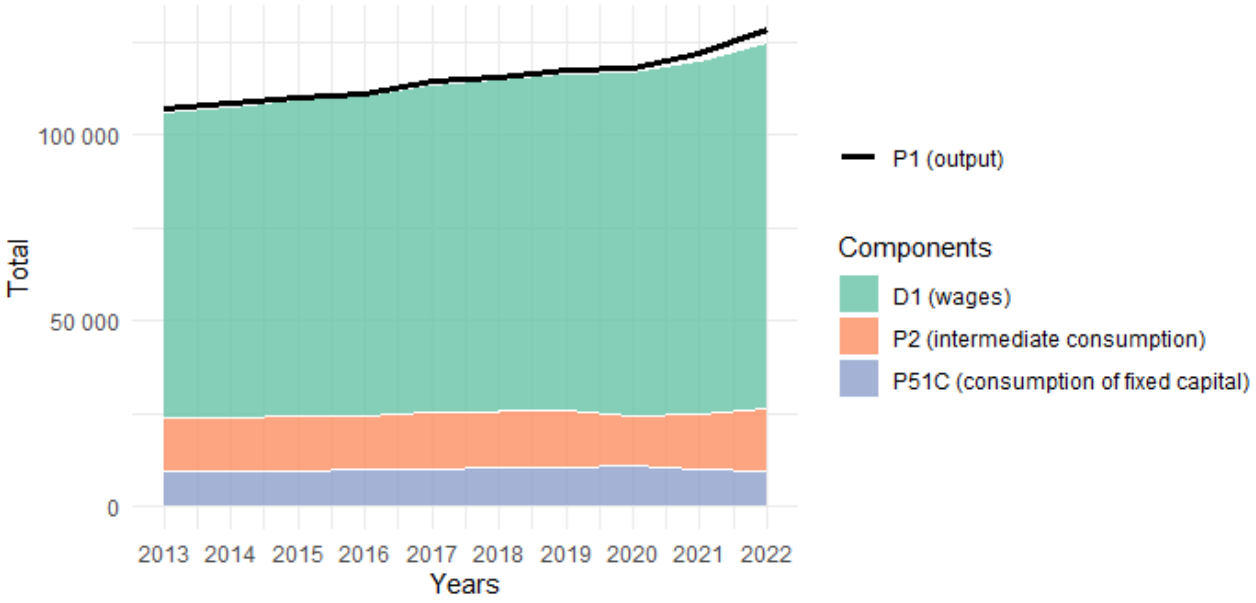
[92] This index is sensitive to the aging of the wage bill: any wage increase driven by seniority is entirely recorded as an increase in worker experience and thus treated as a quality improvement in the input measure, rather than as a pure price effect. This may overstate the quality of inputs if seniority-related wage gains do not fully reflect genuine productivity improvements, which should probably be taken into account when applying this volume–price split method

[93] Panorama statistique des personnels de l'enseignement scolaire 2022-2023, et archives. Although the 2017 salary data were sourced from the 2019–2020 Social Balance Sheets Report, the corresponding employee headcounts for that year were not available. To address this gap, missing values are lineary interpolated.

[94] These two methodological breaks are respectively due to a accounting shift toward an individual approach and the adoption of the so-called "Déclaration Sociale Nominative" which allows employers to pay their social security contributions using company data on their employees

C.2 Figures

Figure 12: Output components for education

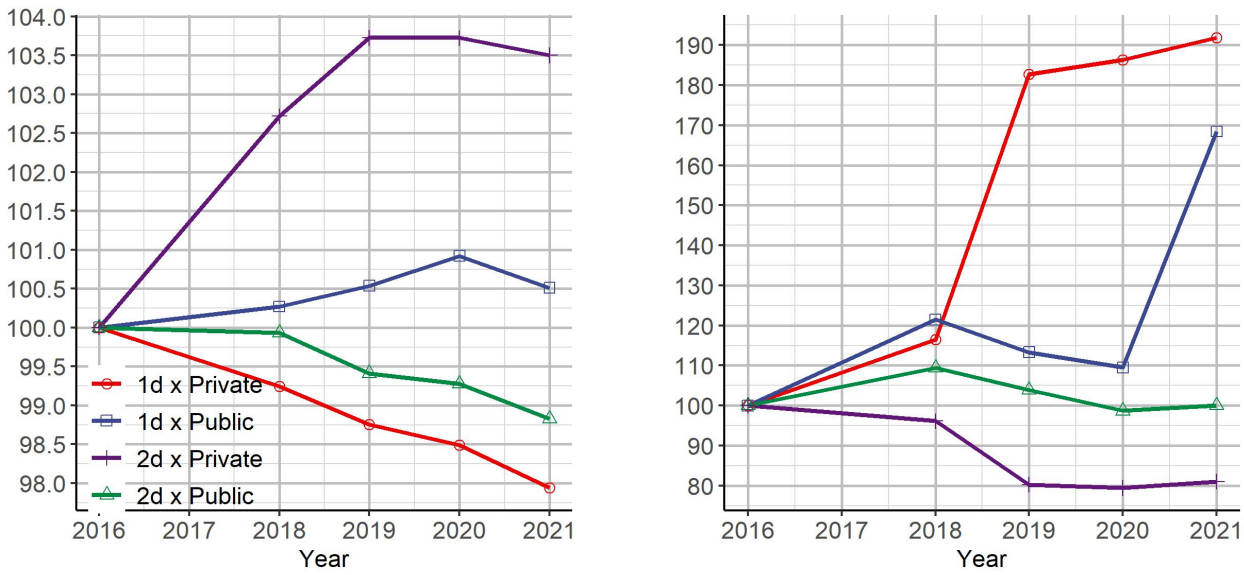


Coverage: Branch 85N (non-market education)
Source: INSEE

Figure 13: Teachers workforce (2016 = 100)

(a) Permanent teachers

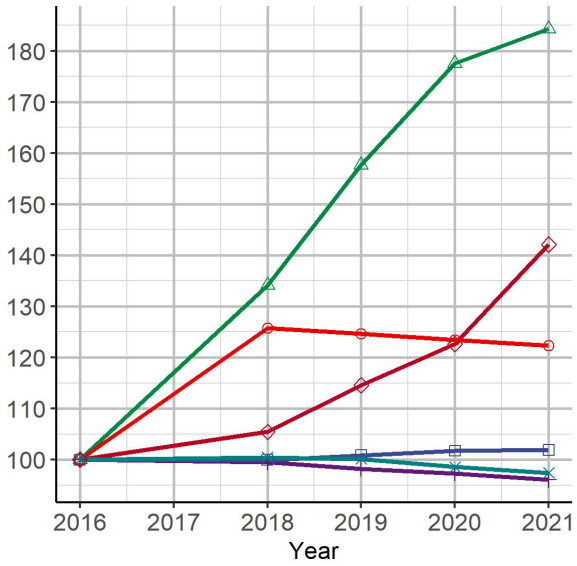
(b) Contract teachers



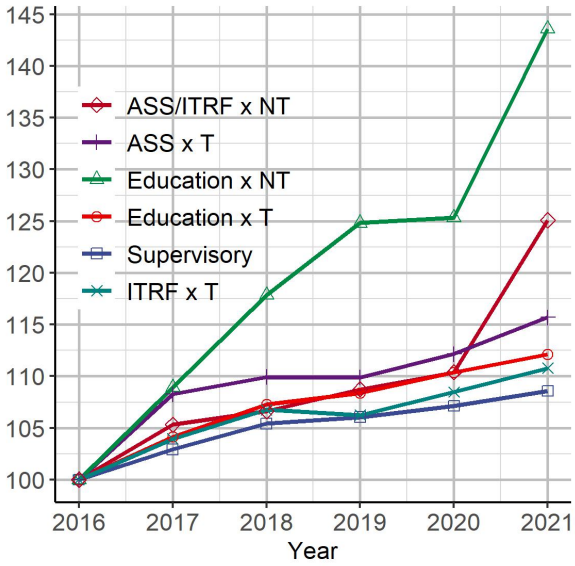
Note : In 2020, the number of FTE permanent teachers in public primary education increased by around 1 % compared with 2016. It's around 10 % for contract teachers.
Coverage : First and second degree education.
Sources : DEPP.

Figure 14: Headcounts and gross wages of non-teaching staff

(a) Headcounts (2016 = 100)



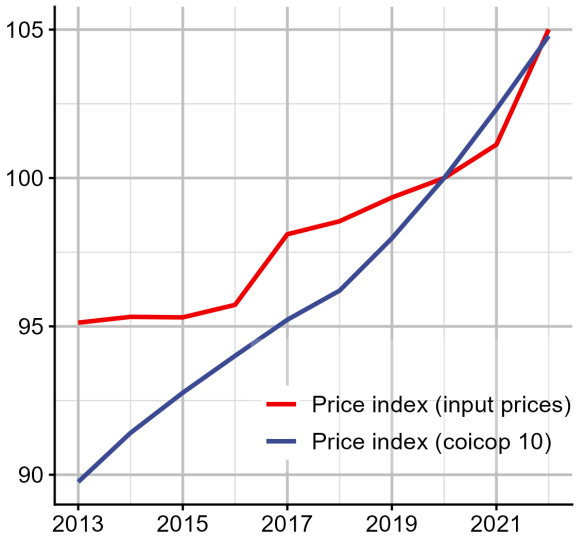
(b) Average gross wage (2016 = 100)



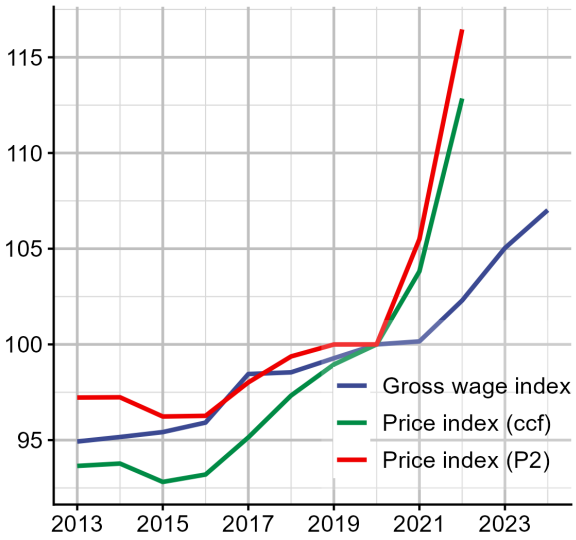
Note : ASS: administrative, social and health staff (*personnels Administratif, Sociaux et de Santé*); ITRF: engineering and technical staff for research and training (*Ingénieurs et personnels Techniques de Recherche et de Formation*); Education: education staff (educational advisors, school psychologists, special needs assistants); Supervisory: school leadership and inspectors. T indicates tenured (permanent) staff and NT non-tenured (non-permanent) staff. Interactions (e.g. ASS × T) refer to occupation-by-tenure status cells. In 2020, the number of non-tenured education staff (in full-time equivalent) increased by around 78 % compared with 2016. Their average gross wage increased by around 25 %.
Coverage : First and second degree education.
Sources : DEPP data.

Figure 15: Prices indices

(a) Input price and coicop price



(b) Price of inputs



Note: The price index based on input prices is computed as a weighted sum of the different components (remuneration of employees, ccf, intermediate consumption).The ccf price index and the intermediate consumption are a weighted price index of the different products consumed by the sector. The price index at the coicop 10 level is the price index of a comparable market service.
Source: INSEE DCN, authors' calculations.

Code	Product	Value
Total		16854.30
PRNA138G_C10A	Meat processing and preservation	1587.50
PRNA138G_F43Z	Specialized construction work	1455.40
PRNA138G_N77Z	Rental and leasing activities	1381.60
PRNA138G_N82Z	Administrative activities	1182.30
PRNA138G_H49B	Other land passenger transport	1027.50

Table 5: Top 5 most consumed products in branch 85N (P2 - intermediary consumption) - Final account 2022

Code	Product	Value
Total		10232.70
PRNA138G_F43Z	Specialized construction work	5657.20
PRNA138G_F41B	Construction of residential and non-residential buildings	2260.00
PRNA138G_F42Z	Civil engineering	1271.00
PRNA138G_J62Z	Computer programming, consultancy and other IT activities	296.50
PRNA138G_C31Z	Furniture manufacturing	216.00

Table 6: Top 5 most consumed products in branch 85N (OPP51C - consumption of fixed capital) - Final account 2022

C.3 Comparing labor volume measures

In the volume of inputs based on input prices, the measure of labor volume is the ratio of total compensation of employees to a wage index based on the index points of civil servants. A key limitation of this wage index is that it excludes bonuses and allowances ("indemnités") that amount to 14 % of teachers' gross wage in 2022, thereby introducing a bias in the resulting volume measure. In contrast, the most common measures of the quantity of labor are headcounts or full-time equivalent (Schreyer and Pilat, 2001). However, such measures fail to account for changes in workforce composition. As a result, they may lead to substantial biases in productivity estimates when the skill mix evolves over time. A direct index of the quantity of labor can overcome the shortcomings of both measurements, assuming that wages reflect productivity.

We compare three measures of labor volume: (i) the measure used in the National Accounts for collective services, defined as compensation of employees deflated by a wage index[95]; (ii) the number of full-time equivalents (FTE); and (iii) a direct Laspeyres volume index constructed from staff counts and net wages. The direct Laspeyres volume index relies on a detailed stratification of the workforce. Teachers are classified by employment status (tenured or contract), sector (public or publicly subsidized private), and level (primary or secondary education). In addition, four categories of non-teaching staff are distinguished[96]. The category-specific weights used in the Laspeyres index are based on average full-time equivalent net wages, which include both allowances and bonuses.

Figure 16 presents those three measures of the quantity of labor for first and second degree education. The three measures follow broadly the same increasing trend over the period 2016-2021. Nevertheless,

[95]For the compensation-based measure, total labor compensation is only available in the National Accounts for the entire education sector, including higher education and apprenticeships. To isolate primary and secondary education, we instead use expenditure on the activity "Teaching" within the sectors "Elementary and pre-elementary schools" and "Middle schools and high schools" from the education thematic account. The wage index corresponds to the average civil service index point, which includes higher education and apprenticeships, and is assumed to provide a reasonable proxy for primary and secondary education.

[96]The four categories of non-teaching staff: student support staff ("vie scolaire"), administrative and technical staff (ITRF), management staff ("encadrement"), and social and health staff (ASS). Student support staff and technical staff are further divided into tenured and contract subcategories.

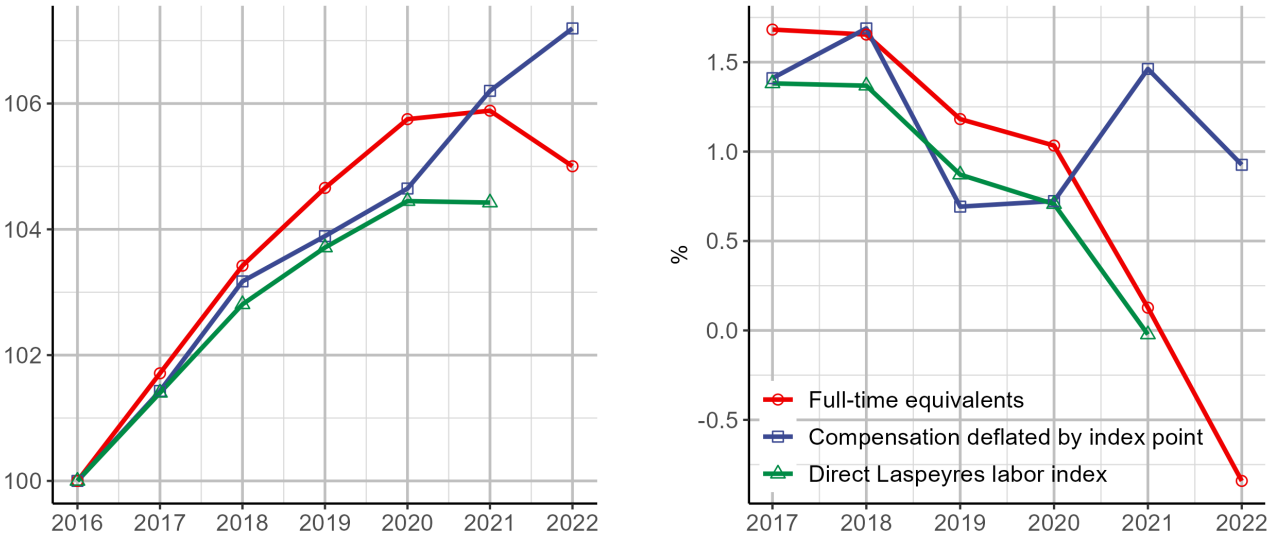
the growth rates of the two chained Laspeyres indexes are lower, because the rise is driven by assistant for pupils with disabilities (AESH), who are in the stratum of non-tenured student support staff, associated to lower wages (around 1,900 euros per month against 2,700 for the average teacher in 2020). Consequently, the chained Laspeyres index assigns them a lower weight to reflect their lower productivity, although the link between wage and productivity is likely weaker in the education sector with intrinsic motivations and government-set salary grids. The compensation deflated by index point grows sharply in 2021 whereas the other measures are stable or decreasing. This is because the Grenelle and computer equipment bonuses created in 2021, as well as the increase in allowances[97], are not accounted for in the index points, thus the corresponding price increases is captured as a volume increase. This underlines the necessity of using precise wage index to compute a labor volume.

The contribution of the different strata of groups of strata can be computed as in Figure 17 in appendix. This shows that student support staff explains most of the increase in labor volume between 2016 and 2021, while the volume of second degree teachers is declining from 2019.

Figure 16: Measures of the labor volume of the education service

(a) Volumes (2016 = 100)

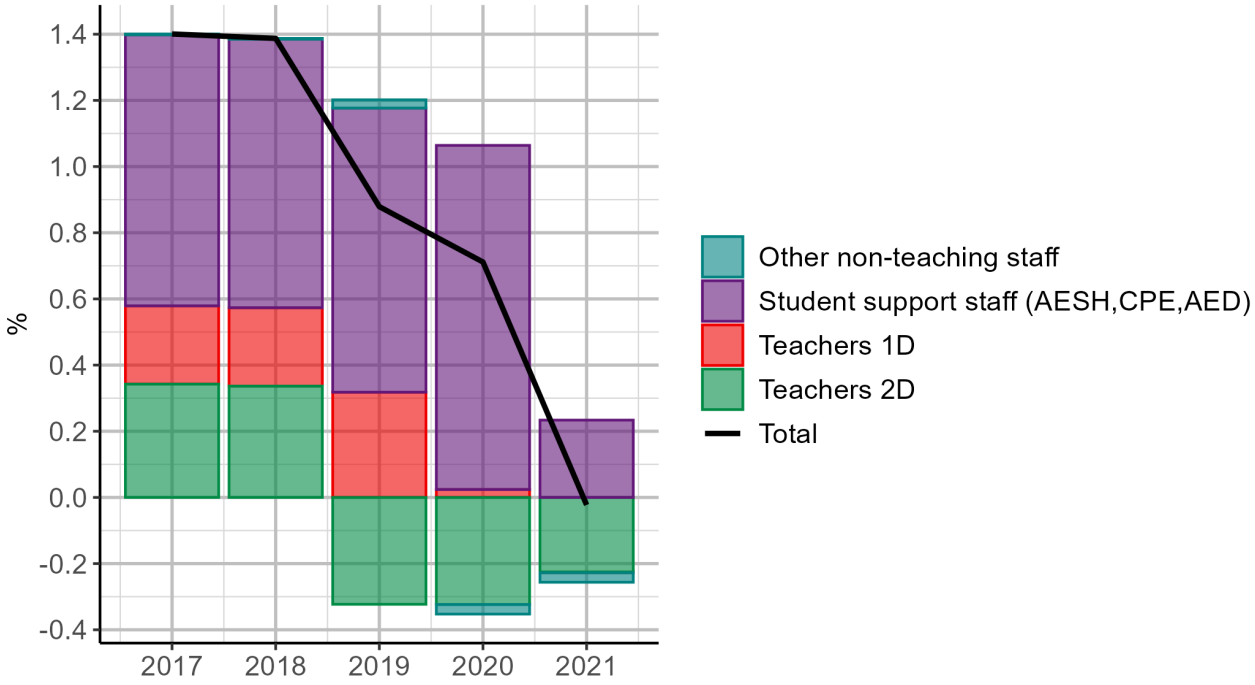
(b) Volume growth rate (%)



Note: Compensations are expense for the activity "Teaching" of the sectors "Elementary and pre-elementary schools" and "Middle schools and high schools" from the thematic account of education. The thematic account is necessary to get the data for first and second degree education only. The wage index used is the average value of the index points including higher education and apprenticeship, which we assume to be a good proxy for first and second degree education only. The direct Laspeyres labor index is computed using headcounts and net wages of teachers and non-teachings staff. The stratification distinguishes teachers depending on their status (tenured or contract teachers) ,their sector (public or subsidies private sector) and their level (first or second degree). The categories of non-teaching staff are: student support staff ("Vie scolaire"), administrative and technical staff (ITRF), management staff ("Encadrement"), social and health staff (ASS). Student support staff and technical staff are further broken down into a tenured and a contract stratum.
Coverage: First and second degree education.
Source: DEPP, authors' calculations.

[97]The computer equipment bonus was 150 euros net for every teachers and the Grenelle bonus was between 500 and 1400 euros net for teachers with less than 20 years of experience. Bonuses were also created for non-teaching staff. Source: https://www.ac-aix-marseille.fr/mesures-de-revalorisation-des-personnels-123274, https://06-site.fsu-snuipp.fr/IMG/pdf/Information-prime-Grenelle.pdf

Figure 17: Contribution to changes in the volume of labor



Note: The Laspeyres quantity index of labor is computed using headcounts and net wages of teachers and non-teachings staff. The stratification distinguishes teachers depending on their status (tenured or contract teachers) ,their sector (public or subsidies private sector) and their level (first or second degree). The categories of non-teaching staff are: student support staff ("Vie scolaire"), administrative and technical staff (ITRF), management staff ("Encadrement"), social and health staff (ASS). Student support staff and technical staff are further broken down into a tenured and a contract stratum. Student support staff include assistants for pupils with disabilities (AESH), heads of pastoral care (CPE), student supervisors (AED).
Coverage: First and second degree education.
Source: DEPP, authors' calculations.

D Stratification adapted to disabled pupils in regular classes

Table 7: Stratification of the education service adapted to the inclusion of students with disabilities in regular classes

Baseline strata	Adapted strata
Pre-elementary	Pre-elementary with disabilities (special classes + regular classes) Pre-elementary without disabilities
Elementary	Elementary with disabilities (special classes + regular classes) Elementary without disabilities
Middle school	Middle school with disabilities in regular classes Middle school without disabilities
General high school	General high school with disabilities in regular classes General high school without disabilities
Vocational high school	Vocational high school with disabilities in regular classes Vocational high school without disabilities
Special secondary	Special secondary (includes special second degree classes)

Note: Special classes include different institutions adapted for pupils with disability (ULIS, UEMA and UEEA). Pupils with disabilities in regular classes attend the same classes as the other pupils, but they can still have assistants (AESH) or special material.

E Output method

E.1 Cost of each strata

Table 8: Share of each level in total education expense of general government, except higher education and apprenticeship

Level	Share of expense in 2013	Share of expense in 2022
Pre-primary	17.0 %	16.6 %
Primary	26.6 %	28.6 %
Lower secondary	27.8 %	27.6 %
Special secondary*	1.9 %	1.5 %
General upper secondary	17.2 %	17.3 %
Vocational upper secondary	9.4 %	8.4 %

Note: *The level "special second degree" includes classes for middle school pupils in acedemic difficulty (Segpa), specialized schools, regional special education schools (EREA), as well as youths in custodial education. The share of expense in each strata $\frac{C_i^t}{C^t} = \frac{c_i^t y_i^t}{\sum_i c_i^t y_i^t}$ with c_i^t the cost per pupil and y_i^t the number of pupil in stratum i in year t is used as weight in the volume index $Q_{output}^t = \sum_i \frac{C_i^t}{C^t} \frac{y_i^t}{y_i^{t-1}}$.
Source: DEPP

F Indicators not retained for quality explicit adjustment

Grade repetition rate According to Schreyer (2010), the French national accounts used the number of students advancing to the next grade as a quality adjustment between 1981 and 2001. Since grade repetition is designed to support struggling students, its proportion can be considered to reflect student performance, particularly for the lowest deciles, and thus the quality of teaching they received. However, using this measure in a time series assumes that the criteria for repeating a grade remain constant over time. French authorities have substantially reduced grade repetition over the past thirty years, following criticism regarding its inefficiency (Crahay, 2004) and high cost (Benhenda and Grenet, 2015). Repetition rates have therefore declined at all levels since 2000, particularly in middle and high school. Due to the lack of comparability over time, we did not retain this indicator for a quality adjustment.

National exam pass rates Results on the national end of high school (*baccalauréat*) and end of middle school (*brevet des collèges*) exams appear to be a natural measure of teaching quality. They are available over a long period, and are taken by nearly all students at the relevant levels. However, they have undergone numerous changes that make them poorly comparable over time (Thomas, 2025): introduction of control exams in 2009, reforms of vocational tracks from 2011 to 2014, continuous assessment for the diploma in 2020 during the Covid pandemic, *baccalauréat* reforms in 2021, and removal of academic correction scales in 2024[98]. They are therefore not suitable for use as consistent quality indicators.

Inspection reports As mentioned by Schreyer (2010), inspection reports can provide useful information on the quality of education services by assessing teaching practices, curriculum implementation, student monitoring, and school management. However, the indicators in French inspection reports are multidimensional, largely subjective, and qualitative. They generally do not follow a metric scale, which is required for volume estimation, and even if they did, it would not be straightforward to determine the appropriate weighting when aggregating different indicators. The information in these reports is also not freely and centrally accessible. For these reasons, we did not use inspection reports as a quality adjustment.

Teacher wages Teaching quality partly depends on a teacher's investment and motivation, which are not directly observable but are linked to the position of the teaching profession within the labor market. This can be proxied by a salary-based index reflecting the attractiveness of the profession, both in terms of recruiting high-quality candidates and retaining the most competent teachers.

Higher teacher wages can contribute to better student outcomes in two ways: by attracting more qualified candidates to the profession, and by providing stronger incentives for teachers to invest effort in their work. Beyond the absolute level of pay, comparing teacher wages to earnings in comparable jobs provides insight into the social standing and overall attractiveness of the profession. A higher relative wage can help retain competent teachers and improve overall teaching quality.

We tested a quality adjustment using teacher wages relative to similar occupations, controlling for the composition of the teaching workforce. This could serve as a measure of occupational attractiveness and competitiveness. We relied on the estimate of Dolton and Gutiérrez (2011) obtained from a cross-country analysis within the OECD, linking higher teacher pay to improved student test scores. We used this macro-level estimate to translate changes in teachers' relative wages into potential effects on students' future earnings within the human capital framework we developed. Between 2010 and 2021, teachers' relative wages fluctuate but display a slight downward trend for secondary education teachers. As a result, the adjusted volume follows similar fluctuations, increasing in years when

[98] https://www.education.gouv.fr/choc-des-savoirs-une-mobilisation-generale-pour-elever-le-niveau-de-notre-ecole-380226

teachers' pay improves relative to comparable occupations and declining when this relative position deteriorate. However, this quality adjustment was not retained because it conflicts with standard national accounting conventions: wages are normally part of price indices, whereas here they would be used to construct a volume-based indicator, making its practical implementation questionable.

Teacher qualifications Another potential quality adjustment could focus on teachers' level of qualifications. Some studies have shown a positive link between teachers' degrees and student performance. For instance, higher academic qualifications and strong evaluation results for trainee teachers tend to improve high school students' outcomes, particularly in mathematics (Clotfelter et al., 2007).

However, detailed data on teachers' qualifications or fields of study are not available in France. Recruitment for permanent teachers is done via competitive exams, requiring a bachelor's or master's degree depending on the type of exam. Beyond these eligibility criteria, we lack information on the proportion of teachers holding a master's or bachelor's, results during training, or the discipline of study. Admission or eligibility thresholds, which could serve as indicators of selectivity, are also not centrally accessible. Similarly, no information exists on the qualifications of contract teachers. In the absence of such data, we did not use a quality adjustment based on teacher qualifications.

G School climate

G.1 Literature review

Academic performance is not the only criterion for assessing the quality of the education service. Indeed, the public education system in France pursues multiple objectives such as reducing inequalities, facilitating professional integration, promoting lifelong learning, and fostering values such as equality and citizenship[99]. The concept of school climate encompasses several of these goals. In France, school climate is assessed through a specific survey conducted by the DEPP: the National School Climate and Victimization Survey[100]. This survey periodically collects responses from primary, middle, and high school students regarding various dimensions of school climate, including school and classroom atmosphere, peer and teacher relationships, perceptions of learning, school facilities, cafeteria conditions, extracurricular activities, and experiences of victimization. It should be noted that responses are also shaped by students' socio-economic backgrounds, which can influence their perceptions of school climate. As a result, reported measures may reflect not only actual school conditions but also differences in social and cultural background.

Ideally, we would like to observe an outcome variable that directly captures all the dimensions pursued by the education system. However, in the absence of such a measure, we restrict our analysis to the effect of school climate on real wages, which likely leads to underestimate the variations in quality linked to school climate, since these directly affect pupils' present and future utility independently of their skills. We leave a more comprehensive evaluation of this issue for future research.

To link school climate to real wages, we combine the effect of school climate on grades with the effect of grades on real wages. Indeed, to our knowledge, no study directly quantifies the impact of school climate on students' human capital. However, the relationship between school climate and academic performance is well-established in the literature. A body of literature shows a positive correlation between school climate and academic performance, as it enhances motivation (Eccles et al., 1993; Goodenow and Grady, 1993) and promotes elements such as cooperative learning, group cohesion, mutual respect, and trust, which facilitate learning (Finnan et al., 2003; Ghaith, 2003). Several studies identify the effect of school climate on academic performance using national data (Shindler et al., 2016; Voight and Hanson, 2017; Tomaszewski et al., 2024) or international data, particularly through PISA

[99] Articles L121-1 to L24-20 of the French Education Code

[100] This survey targets both students and educational staff. Due to data availability, this study focuses solely on students.

surveys (Yu and Jiang, 2022; Baysu et al., 2023). Erdem and Kaya (2024) provide a meta-analysis on the subject, compiling nearly 37 studies. They conclude, based on the existing literature, that a one-standard-deviation improvement in school climate is associated with an average increase of 0.30 standard deviations in academic performance.

G.2 School climate indicator

The National School Climate and Victimization Survey is not conducted annually for all grade levels but provides several waves of results for different educational levels: one for primary schools (2020/2021) [101], three for middle schools (2012/2013, 2016/2017, 2021/2022), and three for high schools (2014/2015, 2017/2018, 2022/2023). The survey consists of a number of questions asked to students, who must validate or reject statements such as "feeling good", "having no absences due to fear" or "learning rather well or very well". The number of statements differs by education level. A higher number of positives answers indicating a more favorable perception of the school environment. Since the data does not provide the full distribution of answers but only an aggregate version, we must make some assumptions detailed in the following paragraph to obtain the mean and standard deviations of scores at each period t.

Approximation of the distribution of climate scores There are 12 statements for medium schools and 17 for high schools. We note $r^t_{i,j}$ the individual score of individual j in stratum i in year t, where the strata i are "Middle school", "Vocational high school" and "general high school". The score $r^t_{i,j}$ lies in $\{0, 1, \dots, 12\}$ or in $\{0, 1, \dots, 17\}$, depending to the education level i. In each case, $R_i \in \{12, 17\}$ denotes the total number of statements. At period t, there are N_i students in level i.

The National School Climate and Victimization Survey do not provide the full distribution $p^t_{r,i} = \frac{1}{N_i} \sum_j \mathbf{1}(r^t_{i,j} = r)$ for $r = 0, \dots, R_i$, but only aggregated proportions by group. For example, we know the proportion of students who validated between 0 and 3 statements at a given time t. Let k run over the aggregated score groups provided by the National School Climate and Victimization Survey. We denote $a_{k,i}$ and $b_{k,i}$ the lower and upper bounds of aggregated group k, $\tilde{p}^t_{k,i} = \sum_{l=a_{k,i}}^{b_{k,i}} p^t_{l,i}$ the share of students in that group, and $m_k = \frac{a_{k,i}+b_{k,i}}{2}$ the midpoint of the interval. Survey's parameters are summarized by the following table :

Table 9: Parameters by School Type

Type of School	a	b	m
Medium Schools	$(0; 4; 7; 10)$	$(3; 6; 9; 12)$	$(1.5; 5; 8; 11)$
High Schools	$(0; 5; 9; 13)$	$(4; 8; 12; 17)$	$(2; 6.5; 10.5; 15)$

Since we do not know the full distribution of $p^t_{r,i}$, we assume a uniform distribution within each group. For all $l \in \{a_{k,i}, a_{k,i} + 1, \dots, b_{k,i}\}$,

$$p^t_{l,i} = \frac{\tilde{p}^t_k}{b_{k,i} - a_{k,i} + 1}$$

Under the intra-group uniform assumption, the mean score is approximated by[102]:

$$\rho^t_i = \sum_{k=1}^{4} \tilde{p}^t_{k,i}\, m_{k,i}$$

[101] Since there is only one observation, we do not adjust primary schools. We also do not adjust secondary special education since it is not covered by the survey

[102] The true mean climate score $\bar{r}^t_i$ for level i can be bounded as: $\rho^t_{i,\min} = \sum_{k=1}^{4} \tilde{p}^t_{k,i}\, a_{k,i} \leq \bar{r}^t_i \leq \rho^t_{i,\max} = \sum_{k=1}^{4} \tilde{p}^t_{k,i}\, b_{k,i}$

Under the same intra-group uniform assumption, the variance of individual scores in level i decomposes into between-group and within-group components[103]:

$$(\sigma_i^t)^2 = \underbrace{\sum_{k=1}^{4} \tilde{p}_{k,i}^t (m_{k,i} - \rho_i^t)^2}_{\text{between-group variance}} + \underbrace{\sum_{k=1}^{4} \tilde{p}_{k,i}^t \cdot \frac{(b_{k,i} - a_{k,i})(b_{k,i} - a_{k,i} + 2)}{12}}_{\text{within-group variance (uniform)}}$$

The National School Climate and Victimization Survey is conducted at irregular intervals. First, we assign each school-year observation to the corresponding civil year $t + 1$ — that is, the calendar year in which the school year ends. This convention reflects the fact that the bulk of the academic year (roughly eight months) falls within year $t + 1$. Second, we interpolate linearly between the resulting civil-year observations.

G.3 Quality adjustment

The literature suggests that improvements in school climate are positively correlated with academic performance, with some studies reporting associations as large as a 0.30 standard-deviation increase in performance for a one-standard-deviation improvement in school climate — though this is likely an upper-bound estimate.To translate this effect into earnings, we combine this relationship with estimates from the literature on the return to academic performance. In particular, Fajeau et al. (2025a) and Hanushek and Woessmann (2006) show that a one-standard-deviation increase in grades is associated with a 9.5% to 30% increase in wages. Interpreting academic performance as an intermediate outcome between school climate and wages, this implies that a one-standard-deviation improvement in school climate leads to an increase in wages between approximately 2.85% (0.30 × 9.5%) and 9% (0.30 × 30%). Therefore, the central calibration (weak adjustment) is :

$$\text{a one s.d. } (\sigma^t) \text{ improvement in school climate } \Rightarrow \text{ +2.85\% in future wages}$$

Then we map these semi-elasticities into an exponential quality adjustment function [104] as :

$$f_i(\rho_i^t) = \exp\left(0.0285 \, \frac{\rho_i^t}{\bar{\sigma}_i}\right)$$

where ρ_i^t is the climate score in level i, and $\bar{\sigma}_i = \frac{1}{T} \sum_t \sigma_i^t$ is the time-average of the cross-sectional standard deviations estimated under the uniform assumption. Only middle school, vocational and general high school strata are adjusted. 0.0285 corresponds to the semi-elasticity which maps the effect of the school climate indicator on future wages. Under the strong adjustment scenario, 0.0285 is replaced by 0.09.

G.4 Results

The curves of the unadjusted and climate-adjusted volumes are indistinguishable on a graph, so we display the gap between their growth rates in Figure 18. The gap is negative throughout the 2016-2022 period which is consistent with the fact that school climate has deteriorated marginally over the sample period, of around −0.032 standard deviations over 10 years for medium schools or 0.018 standard deviations over 9 years for general high schools. The magnitude of the adjustment is, however, negligible. Under the central calibration (weak adjustment), at its largest, the cumulative gap reaches

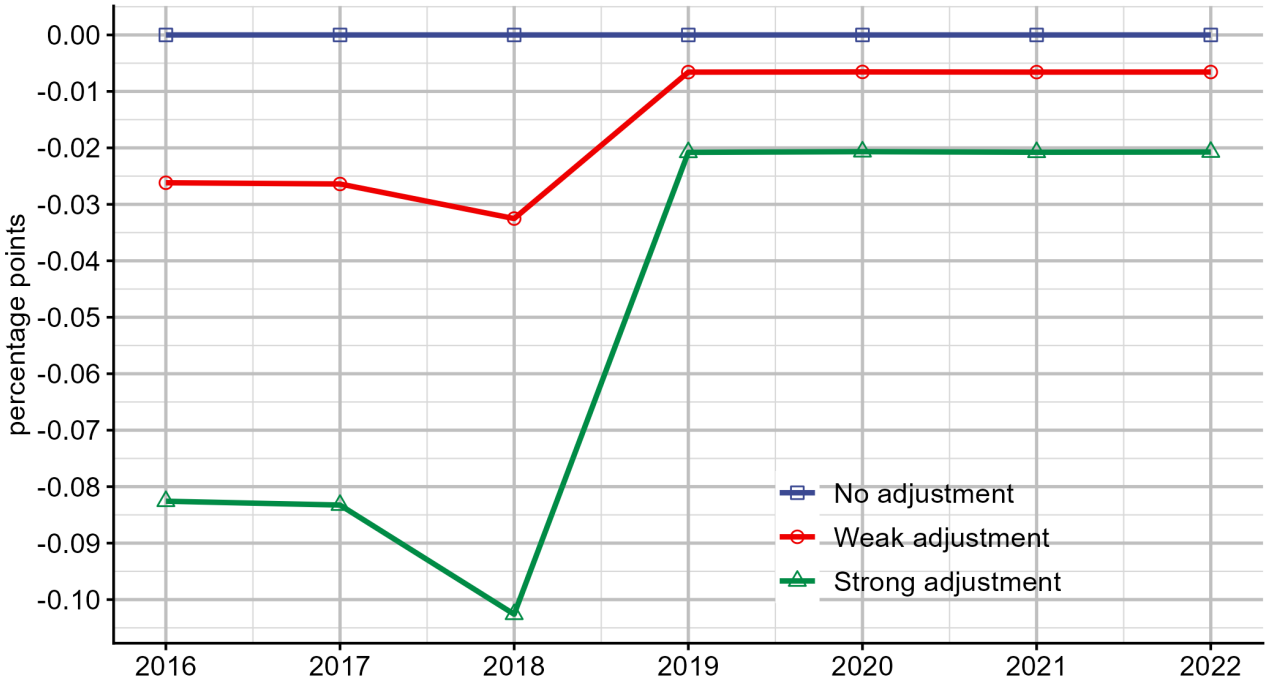
[103] The within-group variance of a group k containing $n_{k,i} = b_{k,i} - a_{k,i} + 1$ consecutive integers under the discrete uniform distribution is $\frac{n_{k,i}^2 - 1}{12} = \frac{(b_{k,i} - a_{k,i})(b_{k,i} - a_{k,i} + 2)}{12}$.

[104] Note that this calibration rests on the assumption that the effect of an aggregate, time-series improvement in school climate is equal to the individual-level effect estimated in the literature

approximately 0.03 percentage point. This finding is a direct consequence of the remarkable stability of the school climate distribution over time: the mean climate score ρ^t declines by less than 0.015 $\bar{\sigma}$ per year across all levels. The 2018 decrease is mainly due to the slight decrease of the climate scores in high school, of around 0.007 standard deviation between 2017 and 2018 [105]. Using the formula derived in section 3.2.3, the contribution of the adjustment to volume changes can be decomposed into three factors, namely the share of adjusted strata in total value, the average change in the school climate index, and the semi-elasticity used for future wages. For the end of the period, that gives :

$$
\begin{aligned}
&\text{weight of adjusted strata } (\approx 0.55)\\
&\times \text{ change in school climate } (\pm 0.01 \text{ s.d. per year})\\
&\times \text{ semi-elasticity } (0.0285)\\
&\approx \pm 0.01 \text{ percentage point})
\end{aligned}
$$

Figure 18: Gap between climate-adjusted volume variations and unadjusted volume variations



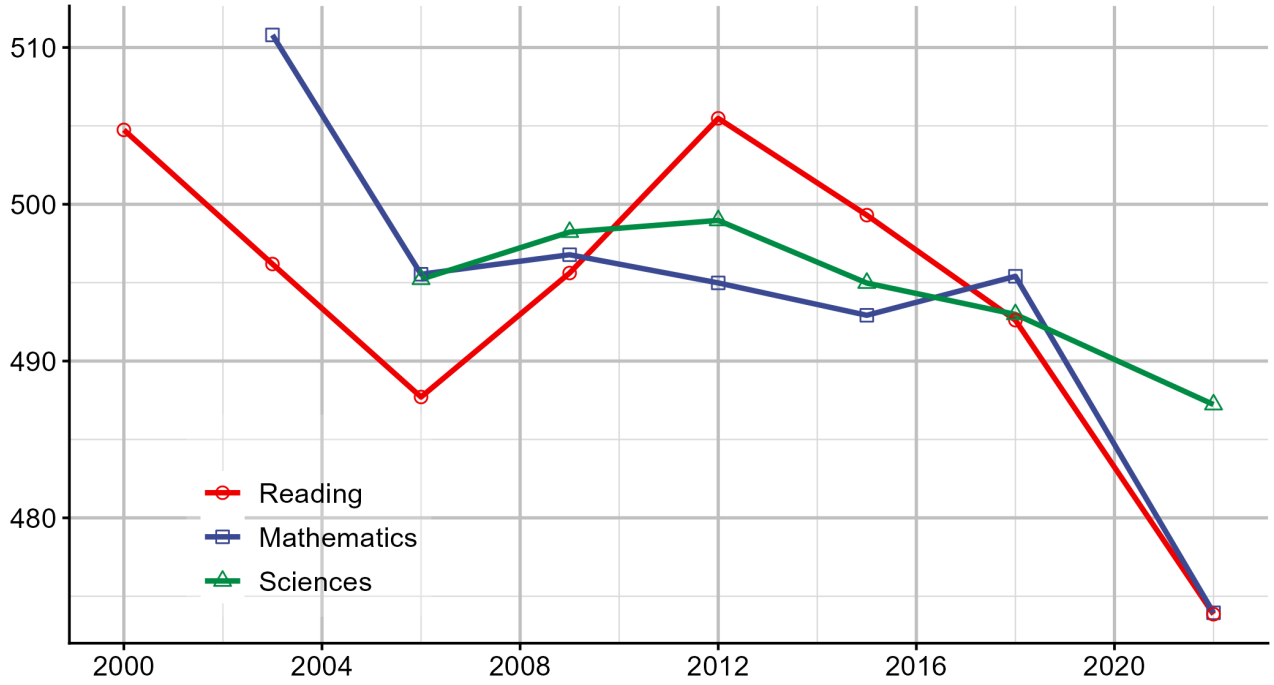
Note: Volume index based on the output method is computed as the sum of relative change in number of students in each stratum, weighted by the cost of the stratum. The adjusted volume is given by $v(y, q) = ye^{\gamma q}$ where q is the mean climate score expressed in units of its time-average cross-sectional standard deviation and γ a parameter measuring the economic returns of a variation of the climate score.
Coverage: medium and high schools
Source: DEPP, authors' calculations.

[105] No clear interpretation was found to explain this variation.

H International tests

H.1 Evolution of PISA scores

Figure 19: Average PISA score over time

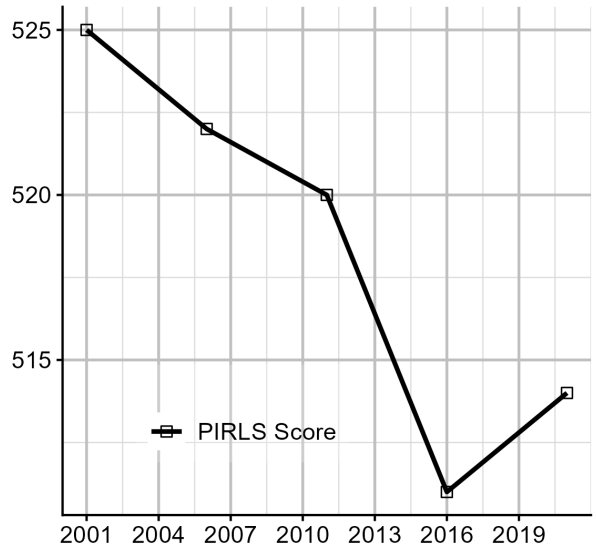


Note: In 2018, the mathematics score at PISA was around 495 for France
Source: OECD

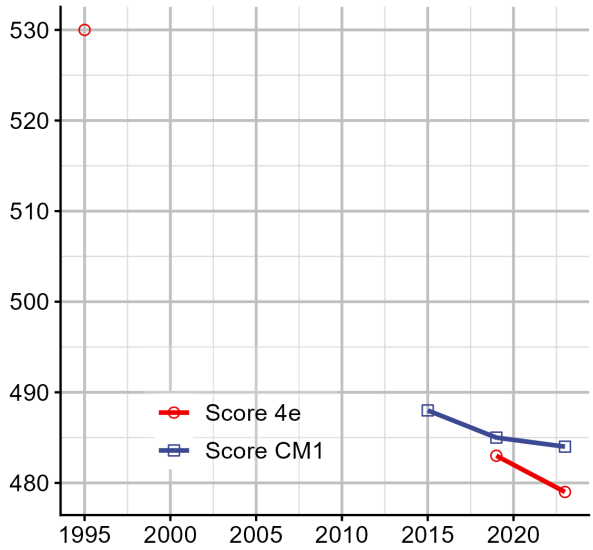
H.2 Evolution of TIMMS and PIRLS scores

Figure 20: Evolution of TIMMS and PIRLS scores

(a) Average PIRLS score



(b) Average TIMMS score

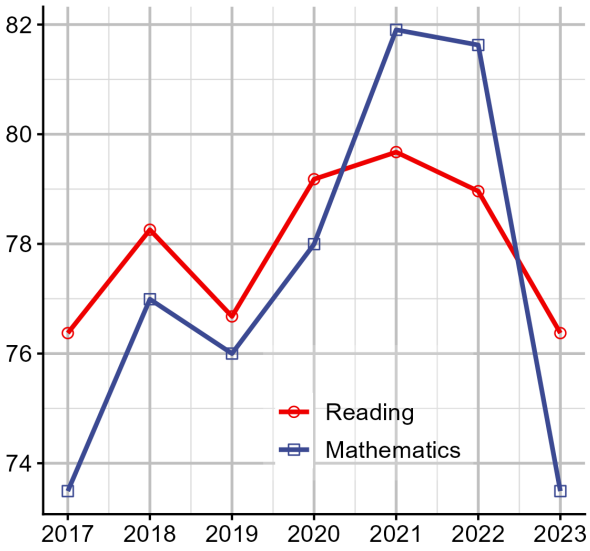


Note: In 2011, the PIRLS score was around 520 for France
Source: IEA

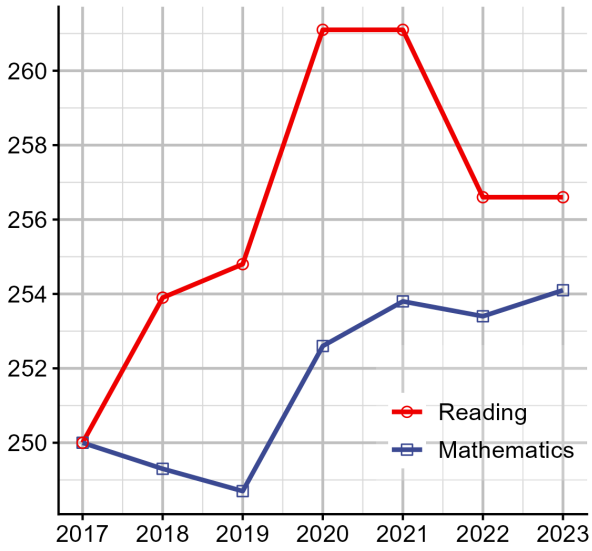
H.3 Evolution of national assessment scores

Figure 21: Evolution of national assessment scores

(a) Grade 1



(b) Grade 6



Note: In 2021, the mathematic score in Mathematics for Grade I was around 83 at the national tests ("évaluation des acquis de CP")
Source: DEPP

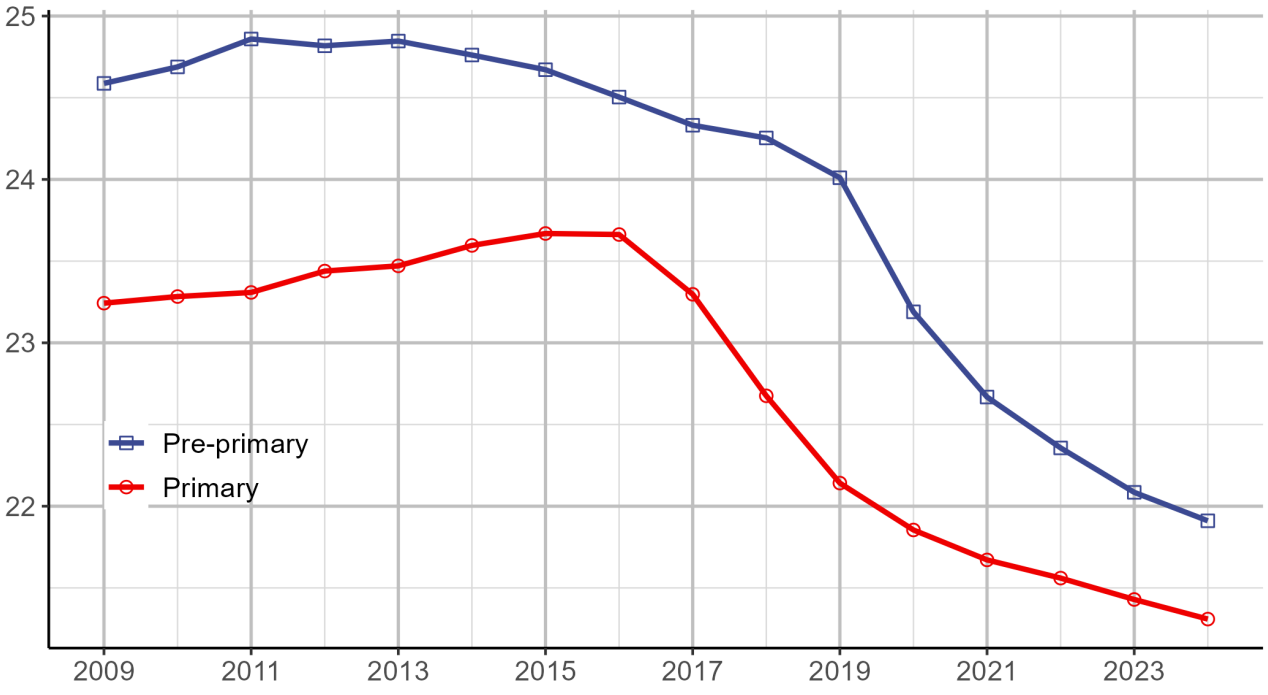
I Class size

I.1 Class size adjustment for secondary education

For lower secondary education, we consider the average effect observed by Piketty and Valdenaire (2006) in last year of middle school and combine it with the effect on wage from Fajeau et al. (2025c) to obtain an elasticity of -0.03 and define $f^{HC-2nd}(q) = q^{-0.03}$, with q the average number of pupil in front of a teacher. An extra pupil in a class decreases the grades by 0,80 % of a standard error in mathematics, 1,1 % in French and 1,6 % in history/geography. The average of those three values is -1.2%, which imply a semi-elasticity of $-0,012 \times 0,095 = -0,11$ % per student in a class. The average class size in lower secondary was $q_{est} = 24,4$ in the year 1998/1999 of the study. This is not exactly equal to the average number of pupil in front of a teacher so we might slightly overestimate the elasticity. We deduce an elasticity $\alpha = -0,0011 \times 24 = -0,026 \approx -0.03$. This adjustment is not robust given the lack of literature on the effect of class size in second degree education: the only other study, is Leuven et al. (2008), who find no significant effect in Norwegian middle schools. For upper secondary grades, we can think that the effect of class size is even smaller because students are more autonomous so the time devoted by the teacher to each student matter less for their success. Moreover, the average number of students in front a teachers used instead of class size for secondary education to take into account group classes probably lead to heterogeneous effects depending on the distribution of groups sizes in the different academic subjects, whereas class size is a clearer variable in first degree.

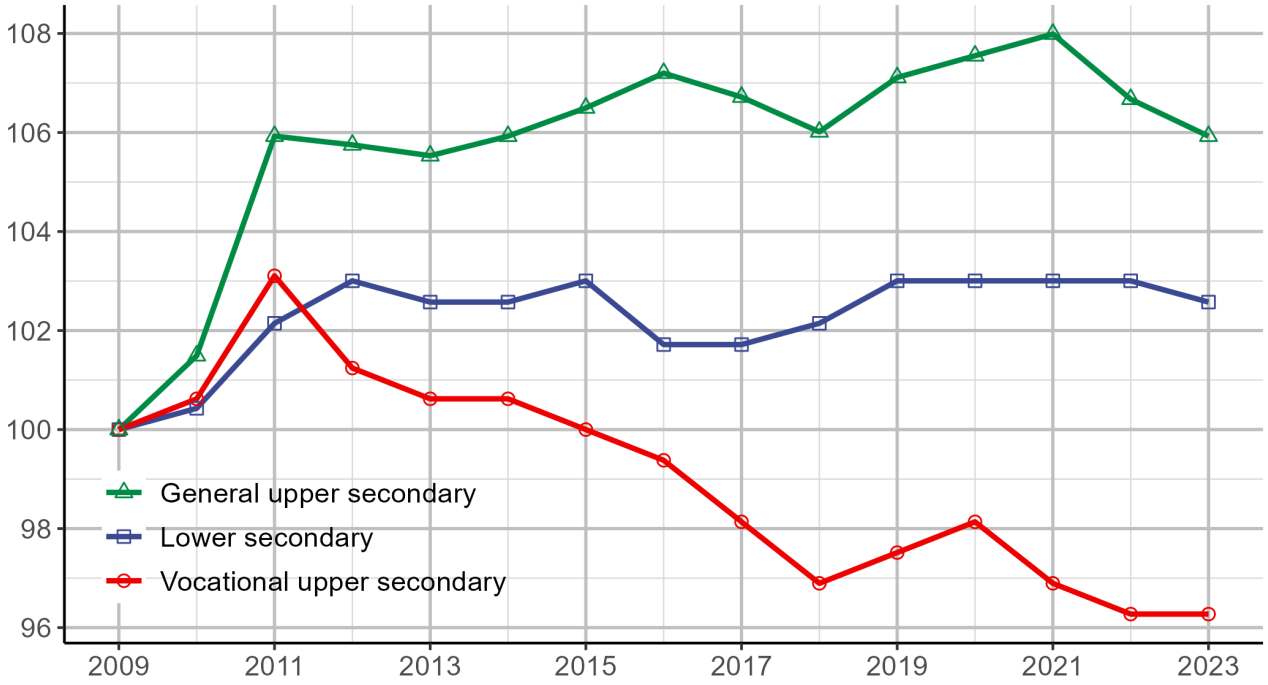
I.2 Evolution of class size

Figure 22: Average class size in first degree education



Sources: DEPP

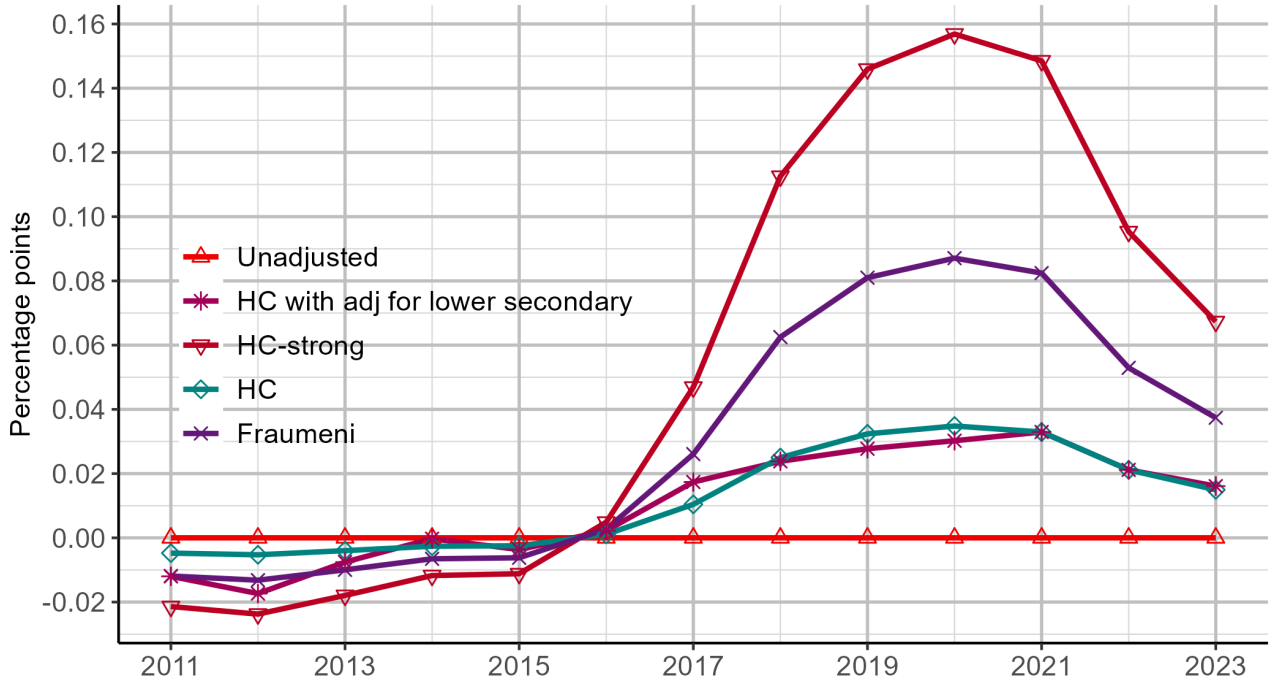
Figure 23: Average class size in second degree education (2009 = 100)



Sources: DEPP

I.3 Adjusted volume growth rates

Figure 24: Gap between growth rate of class size adjusted volume and growth rate of unadjusted volume



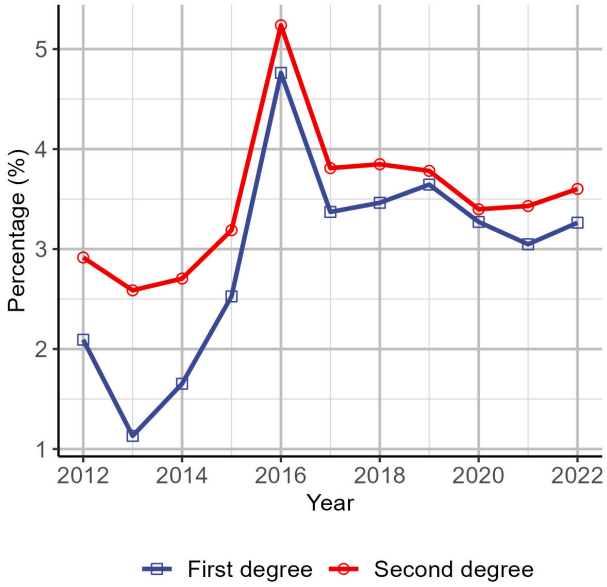
Lecture: In 2019, the growth rate of the chained volume adjusted by class size using Fraumeni et al. (2008) adjustment is 0.081 percentage point higher than the growth rate of the unadjusted volume.
Note: Only pre-primary and primary levels are adjusted for class size, except for the adjustment "HC with adj for lower secondary" where the lower secondary is also adjusted. The adjustments are defined by the adjustments functions $f : q \mapsto f(q)$ where q is class size. $f^{unadjusted}(q) = 1$ correspond to the standard measure using the number of pupils. $f^{HC}(q) = q^{-0.04}$ and $f^{HC-strong}(q) = q^{-0.18}$ are adjustments based the contribution to human capital with constant elasticity, with $f^{HC-strong}$ using the upper bound of the effect of class size on future wage. The adjustment for first degree and lower secondary education uses f^{HC} for pre-primary and primary, and $f^{HC-2nd}(q) = q^{-0.03}$ for the lower secondary strata. $f^{Fraumeni}(q) = q^{-0,1}$ is the adjustment from Fraumeni et al. (2008). Collesi et al. (2007) adjustment is not shown because its growth rate is very different so it would not fit in the y-scale.
Coverage: First and second degree education.
Source: DEPP, Insee, authors' calculations.

J Share of novice teachers

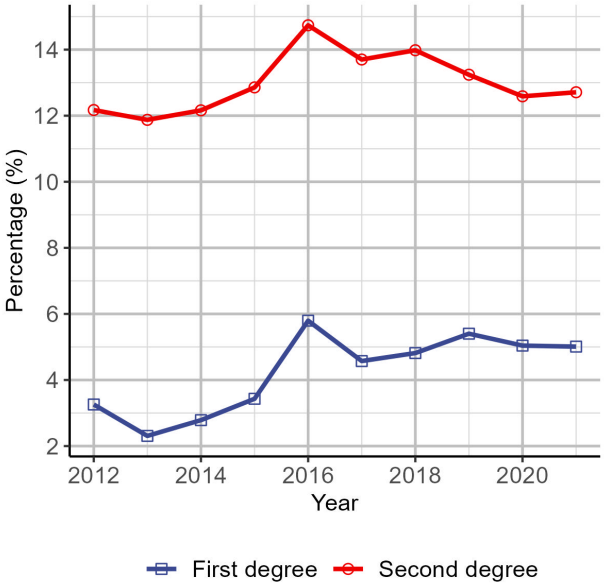
J.1 Evolution of the share of novice teachers

Figure 25: Average share of novice teachers

(a) Novice permanent teachers



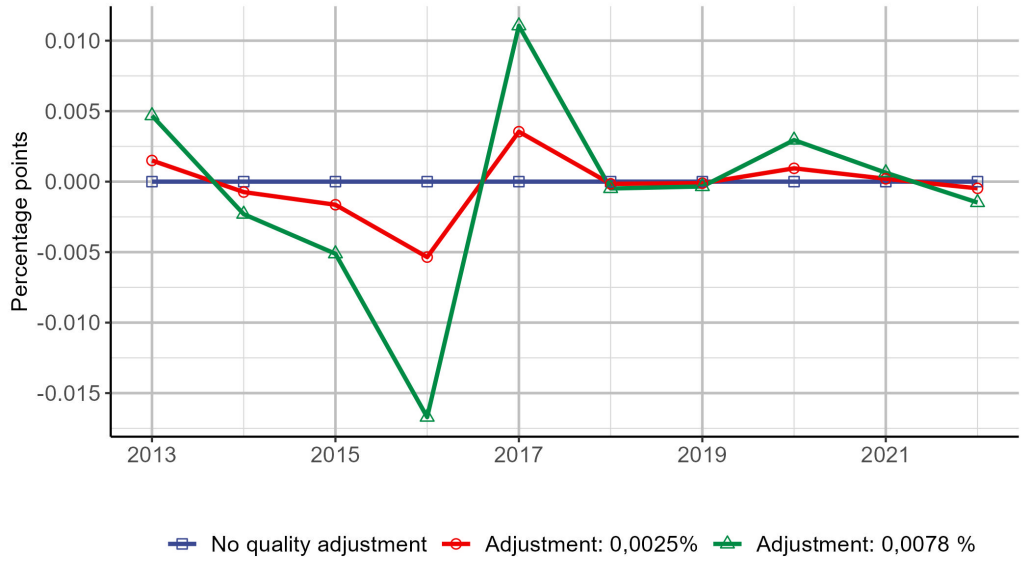
(b) Novice permanent teachers and contract teachers



Note: The left panel shows the evolution of the average share of novice permanent teachers, that is, with zero years of experience, between 2012 and 2022. The right panel shows the evolution of the average share of novice permanent teachers and contract teachers between 2012 and 2021. Between these two dates, in primary education, this share increased by approximately 2 percentage points.
Coverage: First and second degree education.
Source: DEPP, Insee, authors' calculation.

J.2 Results

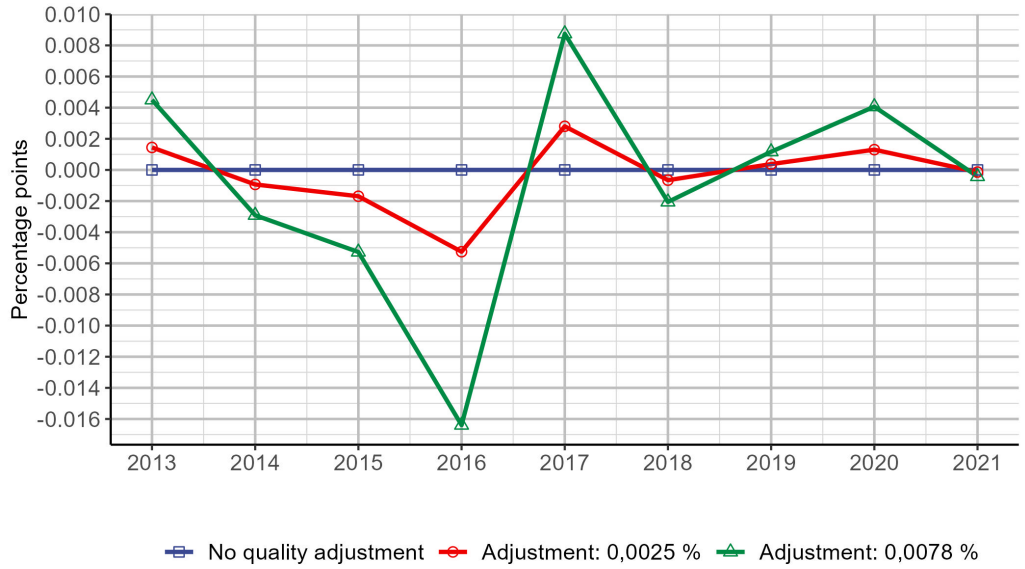
Figure 26: Differences between the growth rate of the adjusted volume and that of the unadjusted volume



Note : In 2017, the difference between the growth rate of the adjusted volume by the share of novice teachers and that of the unadjusted volume is between 0.003 and 0.011 percentage point, for respectively the smallest and the highest adjustment. All strata are quality-adjusted.
Coverage : First and second degree education
Sources : DEPP, Insee, authors' calculations.

J.3 Robustness check : adjustment by the share of novice permanent teachers and contract teachers

Figure 27: Differences between the growth rate of the adjusted volume and that of the unadjusted volume



Note : In 2017, the difference between the growth rate of the adjusted volume by the share of novice teachers and that of the unadjusted volume is between 0.003 and 0.009 percentage point, for respectively the smallest and the highest adjustment. All strata are quality-adjusted.
Coverage : First and second degree education.
Sources : DEPP, Insee, authors' calculations.