Measuring the Effect of Public Bureaucracy on Educational Outcomes in Sierra Leone Using Double Machine Learning *

Kiran Ferrini † Shahin Karami ‡ Kalahan Hughes §

June 15, 2023

Abstract

Educational quality is essential, but remains lacking in many lower and middle income countries. Improving learning outcomes hinges on an efficient and well run government, however properly measuring these qualities is challenging. Can a measurement of bureaucratic quality accurately capture this relationship? We use a World Bank survey of public officials in Sierra Leone to examine this link between bureaucracy and assessment scores. We demonstrate the value of applying a Double Machine Learning methodology to increase precision of causal estimates. Additionally, we provide options for recalculating and reweighing the World Bank indicators for future research. This analysis helps develop a roadmap for future analysis.

1 Introduction

Higher quality of education leads to better and more efficient learning. Educational progress is and should remain a central concern for any government: improved learning outcomes are a tried and true driving force for increasing human capital. A few such positive externalities include higher incomes, productivity, and overall growth (WB, 2017b). Additionally, educated individuals are more likely to competently navigate the job market and find reemployment when unemployed (Riddell and Song, 2011). Ensuring that learning is accessible to all groups of people is essential. Restricted access to education for marginalized groups strengthens economic barriers and emboldens existing inequalities. Increasing learning outcomes for women increases their empowerment and agency, leading to mothers who are well-equipped to raise healthier children (WB, 2017b). However, merely placing children in schools does not lead to increased learning; governments need to know how their policies impact various school inputs and drive learning outcomes.

Improved learning outcomes are crucial to promoting economic growth, economic freedoms, and equality. However, mandating improved learning outcomes is a monumental undertaking. Nations, especially low and middle-income countries, have limited resources requiring governments and education ministries to be well-informed on the efficacy of possible education policies and spending.

*We are grateful to Brian Stacy, Prof. Katya Vasilaky, Prof. Carlos Flores, and Prof. Eduardo Zambrano for their guidance and assistance throughout this project

†e-mail: k.h.ferrini@gmail.com

‡e-mail: skarami141@gmail.com

§e-mail: kalahanhughes@gmail.com

There is an extensive literature in development economics resources devoted to studying the effect of primary practices—direct school inputs such as physical resources, teachers, and school management —on learning outcomes; however, less is known regarding the impact of educational governance on learning outcomes. Educational ministries and governments dictate the policy that drives primary practices. Because of this, governing school boards must be effective for educational investment to occur. Thus, we have employed various econometric strategies to understand this relationship better.

We use a specialized survey of government and administration officials undertaken by the World Bank, including test scores and other learning outcomes for specific schools and districts. The World Bank survey includes data from Sierra Leone; 232 complete school data cases then got linked to their corresponding public official's responses. The data reflects public officials' views on the state of their education and bureaucratic system, aggregated into four indices: National Learning Goals (NLG), Mandates and Accountability (MA), Quality of Bureaucracy (QOB), and Impartial Decision-Making (IDM). In these field surveys, public officials and school administrators were asked qualitative questions. Using this data is beneficial as the enumeration will help shield the analysis from response bias. However, enumeration introduces a possible enumeration effect that could bias the data. These questions could be considered sensitive: officials asked provided honest outlines of bureaucratic operations and efficiencies. This analysis and data aim to provide policymakers in low and middle-income countries with decision-making tools for improving learning outcomes.

In order to measure the quality of each administrative office, we first conduct standard Ordinary Least Squares (OLS) regression analysis. However, unobservable factors or complex interaction terms likely drive public officials' opinions and allocation of resources. We next build upon the OLS analysis with an Instrumental Variable (IV) model to account for unobservables confounders. We then employ a Double Machine Learning (DML) approach to enhance the power of the causal estimates of bureaucratic quality on learning outcomes through regularization. Finally, we conducted a robustness check using three distinct data aggregating methods. These methods included: weighted aggregation using the arithmetic mean, standard aggregation using the quasi-arithmetic mean, and weighted aggregation using the quasi-arithmetic mean.

Comprehensively, we can recommend alternative methods for aggregating educational survey and indicator data. We find evidence that alternative weighting techniques could provide more meaningful estimates of bureaucratic efficiency. Our results also show that DML methods are ideal for deriving estimates of bureaucratic efficiency —as the regularization mechanisms used in DML help to identify relevant covariates without the structural assumptions of other approaches. DML allows for more precise and non-linear causal estimates than traditional methods like OLS and IV approaches.

The results we discuss in this paper contribute to the literature in two ways. First, our results contribute to the growing literature on global initiatives for education access. In 2000, the United Nations established the Millennium Development Goals (MDGs), with the aim to reduce global poverty, improve education access, among other objectives. One of the critical targets of the MDGs was to achieve universal primary education (UN, 2000). By the end of 2015, Progress toward this goal was seen as a success. With significant increases in primary school enrollment rates in developing countries, concerns about the quality of education students receive in these countries also increased. Concerns such as these led to the adoption of the new Sustainable Development Goals

(SDGs) in 2015, which placed a greater emphasis on improving the quality of education (UN, 2015).

Over the last 25 years, a growing body of literature has examined methods to increase the quality of education and student participation in the classroom for the explicit goal of greater learning outcomes. For example, studies have shown the benefits of hiring extra short-term contract teachers on educational outcomes versus hiring an extra civil service teacher. Benefits include reduced teacher absenteeism, increases in pupil test scores, and in some instances, a cost-efficient alternative to current practices (Bau and Das, 2017; Duflo et al., 2015b; Muralidharan and Sundararaman, 2013). Other studies highlight the benefits of computer-assisted education programs in Banerjee et al. (2007); Lai et al. (2015). However, studies such as Banerjee et al. (2007) found that providing free textbooks to students does not raise the average test scores. Banerjee et al. (2010) highlight how the success of an educational program is contingent on whether government bureaucrats or nongovernmental organizations implement them.

These new SDGs place an increased emphasis on improving the quality of education, with a particular focus on ensuring access to quality education for all. In order to achieve this goal, it is essential to understand and address bureaucratic inefficiencies and improve the quality of education delivery systems in developing countries. Therefore, further research is needed to identify how government bureaucracy influences learning outcomes and develop effective policies and interventions to address bureaucratic inefficiencies in the education sector. By developing an understanding of the quality of bureaucracy, developing countries can take a step towards achieving the SDGs and ensuring that all children receive a quality education that prepares them for future challenges.

Second, our results contribute to the literature aimed at improving education quality. (Asongu and Odhiambo, 2019; Banerjee et al., 2007; Beaman et al., 2012; Duflo et al., 2015a, 2021; Hill and Chalaux, 2011; Kremer et al., 2013; Miguel and Kremer, 2004; Murtin, 2013; Smidova, 2019). WB (2017a) notes the importance of bureaucratic quality as a complement to factors such as school infrastructure and teacher training in improving education quality and equity. There is literature that supports the theory of government decentralization leading to economic growth (Barankay and Lockwood, 2007; Clark, 2009; Faguet and Sánchez, 2014; Galiani et al., 2008). These papers provide a deeper understanding of the bureaucracy within the country of interest. However, they do not deeply discuss the bureaucracy's efficiency or quality concerning educational outcomes.

The effectiveness of teachers and staff is known to influence student learning outcomes, but research highlights the critical role that education administrators play in mediating this relationship (Hallinger and Heck, 1996; Leithwood et al., 2008; Robinson et al., 2007). While some studies have found that increasing the training of administrative officials has minimal impact on learning outcomes (Haller et al., 1994), Hallinger and Heck (2011) have argued that school leadership indirectly exerts a measurable effect on student learning. Robinson et al. (2007) similarly reports a moderately strong effect of school leadership on student outcomes, with "promoting and participating in teacher learning and development" having the largest effect and "planning, coordinating, and evaluating teaching curriculum" having a moderately large effect. Meanwhile, other dimensions identified have a minor impact on student learning. Survey data can measure the dimensions of administrative impact, as it allows for a better understanding of the complex relationships between administrators, teachers, and student outcomes. (Heck et al., 2000; Waugh, 2002) The presence of corruption is another dimension that previous studies do not mention but is critical to consider when analyzing the effectiveness of school administration.

Having accurate measures of the effectiveness of school administration is also necessary to reduce one of the most detrimental sources of poor administrative quality: corruption. Research has shown that corruption within education systems negatively correlates with a country's income level and that the prevalence of corruption in lower-income countries can lead to low school enrollment and high dropout rates. (Hallak and Poisson, 2005). In Nigeria, corruption has been found to result in reduced funds for administrative functions, inadequate infrastructure, a shortage of academic staff, poor quality of education, resource wastage, increased administrative costs, and diminished perceptions of universities in the international community (Jacob et al., 2021). However, addressing corruption has become more complex in middle-income countries. For example, a study found that effectively addressing corruption in Romanian Baccalaureate exams unequally impacted low-income students, leading to lower test scores and university acceptance rates (Borcan et al., 2017). These findings suggest that corruption is worse in lower-income countries, and addressing it has a greater potential to increase overall welfare by eliminating this potential externality.

Increasing insights into the benefits and efficiency gains of increasing the quality of bureaucracy begs the question: why have there not been more empirical links to educational outcomes and the effectiveness of educational policy or interventions? The core issue at hand is measurement. Despite numerous indices and indicators for bureaucratic efficiency, there is no consensus on classification, and therefore there are no consistent estimates for levels of government efficiency (Heywood, 2014; Olken and Pande, 2012). Beyond the estimation problems, any empirical results are further obfuscated by the wide range of index inputs and thus not actionable. A solution to this problem is through primary surveys, which attempt to measure the bureaucratic inefficiencies of interest directly. Furthermore, this report is interested in the specific attributes of bureaucracy that affect learning outcomes. A recent example of this methodology is applied in Pakistan (Callen et al., 2023) with a survey of health officials and clinic doctors to measure bureaucratic efficiency and link it to doctor absenteeism. This report utilizes a similar methodology to extend the existing literature on learning outcomes, bureaucratic efficiency, and the links between the two.

In the next section, we provide historical background on education in Sierra Leone. In the third section, we discuss the data used in this report in depth and ethical and bias implication the data may have. We describe our empirical model of public official survey response effecting learning outcome in the fourth section, and our empirical approaches to estimating this relationship. The fifth section presents our main results, discussions around each empirical approach used, and secondary results from three alternative indices as a robustness check. We present concluding remarks in the final section of this paper.

2 Education in Sierra Leone

The history of Sierra Leone is autocracy and bloodshed. In 1961 it declared its Independence from the United Kingdom and became a civilian democracy; then, Siaka Stevens entered politics in 1968 as Prime Minister of Sierra Leone. When Stevens stepped down from leadership in 1985, Sierra Leone was a one-party state (Jang, 2012). During and after Stevens' rule saw the rise of corruption in virtually every sector of Sierra Leone's government. Corruption grew to the level that saw Sierra Leone's government slash the budgets of the health and education sector (Banya, 1991). Prior to the start of the war, roughly 55 percent of all primary school ages and less than 30 percent of secondary school age children were registered in school (Banya, 1993; Jang, 2012; Keen, 2003; Leone, 2001; Reno, 2003)

The 11-year civil war began in March 1991 when the Rebel United Front (RUF) began to cease diamond-rich portions of eastern and southern Sierra Leone. By the end of 1991, the Sierra Leone Army (SLA) had engaged the RUF, and the conflict escalated into civil war. Between the RUF and SLA, conservative estimates say that 12,000 - 20,000 child soldiers were used, with roughly half of the forces the RUF used being child soldiers (McKay, 2005; Peters and Richards, 1998). Between the start of the conflict and its end in January 2002, nearly 50 thousand people perished, and nearly 5 million were displaced (Taylor, 2003). During the civil war, numerous schools were targeted for destruction, so much so that by the late 1990s, an estimated 70% of all children left in Sierra Leone had no access to education Maclure and Denov (2009). During this time, women were affected the most severely. Estimates show that anywhere from 215,000 and 275,000 women were sexually assaulted during this time. By the conclusion of the conflict, only 20% of eligible girls were enrolled in primary school Amowitz et al. (2002); Maclure and Denov (2009).

The post-war period has seen a high demand for quality education, although there have been difficulties in achieving this given the generation of children involved in the conflict (Wang, 2007). In the years following the conflict, the gross completion rate for a primary school in Sierra Leone was only 65%. Many students were also behind in their education due to a late start, repetition, or interrupted schooling (partly caused by time spent as child soldiers) (Betancourt et al., 2008; Wang, 2007). Two government programs have attempted to meet the post-war educational needs of over-age youth: Complementary Rapid Education for Primary Schools (CREPS) and Rapid Response Education Program (RREP) (Wang, 2007). Both of these programs were successful first steps in increasing post-war school enrollment. However, there was evidence of a lower quality of education at schools funded through these programs (Glennerster et al., 2006). The Education Act of 2004 removed all fees related to schooling and made education compulsory through secondary school. Removing fees caused a massive increase in primary school enrollment, which doubled from 660,000 in the 2001—02 school year to 1.3 million in the 2004—05 school year (Wang, 2007). There is progress to increase school enrollment, education quality, and gender equity within education (Hinton, 2009; Maclure and Denov, 2009; McDermott and Allen, 2015). However, more will be needed to maintain and build upon Sierra Leone's progress post-civil war.

3 Data

The World Bank collected the data through several field surveys where enumerators interviewed public officials and schools. Enumerators asked a series of 52 questions public officials. We then averaged them into four primary indicators: NLG, MA, QOB, and IDM. The overall score (1 to 5) is the average of all relevant questions. The NLG indicator comprises 12 questions that measure how each public office operates concerning overall goal targets, progress monitoring, internal incentives, and community engagement. The MA indicator includes nine questions that aim to measure the clearness or coherence of overall goals, how transparent or public the goals are, and how accountable public officials are. The QOB indicator is an aggregation of 12 total questions that measure officials' competency, the work environment's efficiency, the degree to which the office is merit-based, and general morals or attitudes. The IDM indicator is composed of 12 questions that aim to measure the influence of politics on office policies and the effectiveness of unions in both schools and bureaucracy. Then we aggregated all four main indicators (NLG, MA, QOB, and IDM) questions into an overall measurement: Bureaucratic Efficiency (BE).

Linking effectiveness to outcomes requires a consistent measure of school performance. Data collection included surveys for principals and teachers and direct observation from the surveyor. Student knowledge is the average of all the student's scores on math and language assessments given to a randomly selected class of fourth graders by World Bank enumerators. Student proficiency and the 25 included covariates, an aggregation of other survey questions, and enumerator observations have explicit variable descriptions in the code documentation. Each school has been linked with its corresponding four indicators by public officials. Table 1 summarizes our data across schools in Sierra Leone with 232 observations present.

3.1 Ethical and Bias Implications

The findings of this report may have potential policy implications for education administration, but any actions taken to address the identified issues must use caution. It is essential to consider that policy implementation to correct undesirable educational behaviors can disproportionately affect disadvantaged groups harshly, as studies have shown (Borcan et al., 2017). Education policy can disproportionately affect lower-income groups due to limited access to resources, inadequate infrastructure, and reduced support systems, creating barriers that hinder their educational opportunities. Additionally, implementing policies without considering lower-income communities' unique needs and challenges can further widen the educational achievement gap and perpetuate socioeconomic disparities. The information in this report is most relevant to school administrators in the countries where the data has context. These administrators possess a nuanced understanding of the contextual factors that may have influenced the findings, and their voices should carry the most weight in discussions about potential policy changes. One of the key challenges in collecting data is selection bias which occurs when the sample used for research or analysis is not representative of the entire population. Bias can result in inaccurate or misleading conclusions about the overall population.

For instance, in our data set, we only examine survey data from public officials who may only have knowledge of their own experiences and may not need to be aware of students' actual conditions. This selection bias can result in a lack of adequate understanding of the realities in the classroom, including issues related to teaching, curriculum, and other educational problems. For example, our data probably represent schools and school officials with greater access to funding. As Goldstein and Dabla-Norris (2004) shows, funding for public education is often skewed towards the rich because the rich are more effective at appropriating funding despite the education being public. For this reason, any policy implications arising from this study will not reflect the circumstances of more marginalized populations.

Furthermore, human bias may also be present among public officials. They may have, consciously or unconsciously, portrayed the effectiveness of their administration more positively or negatively. These officials may exaggerate certain conditions, such as the quality of education, to highlight their political party's success (Akhtari et al., 2022), leading to a distortion of the actual conditions in the education system. For example, public officials could decline to answer survey questions,

Table 1: Summary Stats for Sierra Leone Schools Controls and Student Knowledge (N = 232)

Variable	Mean	Std. Dev.	Min	Max
Absence Rate	29	24	0	100
Student Attendance	65	26	5.8	100
Students Enrolled	323	203	72	2006
Content Proficiency	11	22	0	100
ECD Student Proficiency	30	40	0	100
Infrastructure	1.8	0.94	0.14	4.6
Teach Score	2.4	0.34	1.3	3.3
Operational Management	4	0.51	2.5	5
Instructional Leadership	3.7	0.94	1.5	5
Principal Knowledge Score	3.4	1	1	5
Principal Management	4	0.61	1.3	5
Teacher Attraction	3.6	0.42	2.5	4.7
Teacher Selection Deployment	3.2	0.86	2	5
Teacher Support	3.4	0.7	1.8	5
Teaching Evaluation	3.9	0.85	1	5
Teacher Monitoring	2.4	0.8	1	4
Intrinsic Motivation	3.8	0.49	2.5	4.9
Standards Monitoring	4.3	0.79	1	5
SCH Monitoring	3	0.71	1	3.9
SCH Management Clarity	4.6	1.2	1	5
SCH Management Attraction	3.5	1.2	1	5
SCH Selection Deployment	4.4	0.62	1	5
SCH Support	4	0.98	1	4.8
Principal Evaluation	4.6	0.64	2	5
Light GDP	0.93	2.6	0	26
Student Knowledge	37	16	9	92

[1]The enumerator collected these variables through direct observations, surveys given to schools and public officials, or knowledge assessments. Variables obtained through direct observation are absence rate, student attendance, students enrolled, and infrastructure. Variables obtained through surveys are: teach score, operational management, instructional leadership, principal management, teacher attraction, teacher selection deployment, teacher support, teacher evaluation, teacher monitoring, intrinsic motivation, standards monitoring, SCH monitoring, SCH management clarity, SCH management attraction, SCH selection deployment, SCH support, principal evaluation. Variables obtained through knowledge assessments are content proficiency, student proficiency, principal knowledge score, and student knowledge.
[2]Light GDP is an informal measure of economic activity using light.
[3]See the code documentation for explicit variable definitions.

potentially due to concerns about political backlash, fear of consequences, or misalignment of interests, introducing selection bias into the sample data, resulting in a lack of accurate representation and spurious regression results.

Another possible instance of selection bias that could be present in our data comes from the way that the World Bank administered the surveys. Due to time and budget constraints, they could not get to every school district, indicating that the sample could be biased toward more available schools. It is important to note that other biases may also be present, and caution should be used when interpreting and utilizing the specific effect estimates of this report. However, our empirical strategy does attempt to address some of these biases, which we describe in the next section on empirical estimation.

4 Problem & Estimation Procedure

The existing literature on education focuses on improving educational outcomes, but few studies do so from the perspective of improving the educational bureaucracy. Measurement of educational bureaucracy is rarely captured. Although many metrics and gauges are available to assess bureaucratic effectiveness, there is a lack of agreement regarding categorization, resulting in inconsistent evaluations of government efficiency (Heywood, 2014; Olken and Pande, 2012). Identifying methods for the government to improve learning outcomes poses a challenge due to the influence of multiple factors on students' academic performance, including personal constraints related to limited financial resources. In addition, estimating causal impact rather than correlational associations between educational bureaucracy and educational outcomes poses a greater challenge.

4.1 Estimation Strategies

In Figure 1, we posit the structural model that relates bureaucratic efficiency and educational outcomes in the style of directed acyclic graphs (DAGs) (Cinelli et al., 2020; Pearl, 2009). The Public Official Indicator refers to one of the five indicators for public officials' opinions: BE, NLG, MA, QOB, and IDM. Student Knowledge measures the sum of the arithmetic mean scores for students in a fourth-grade class on a literacy and mathematics assessment administered by the World Bank. To our knowledge, this report is the first to use this survey data to link the relationship between bureaucratic opinion and Student Knowledge. To refrain from passing judgment on the role that each covariate has within the estimation, we control for the twenty-five covariates observed within the school survey, along with Light GDP and External Infrastructure denoted by the green circles in Figure 1. More importantly, the broad range of covariates within the survey data may contain undiscovered endogenous effects on the Public Official Indicators and Student Knowledge. In this report, we are concerned with estimating the overall effect that public officials have on learning outcomes, shown as the large direct path from bureaucratic efficiency in Figure 1. To reduce potential estimation bias, we employ three strategies: Ordinary Least Squares (OLS) with controls, an Instrumental Variable (IV), and Double Machine Learning (DML).

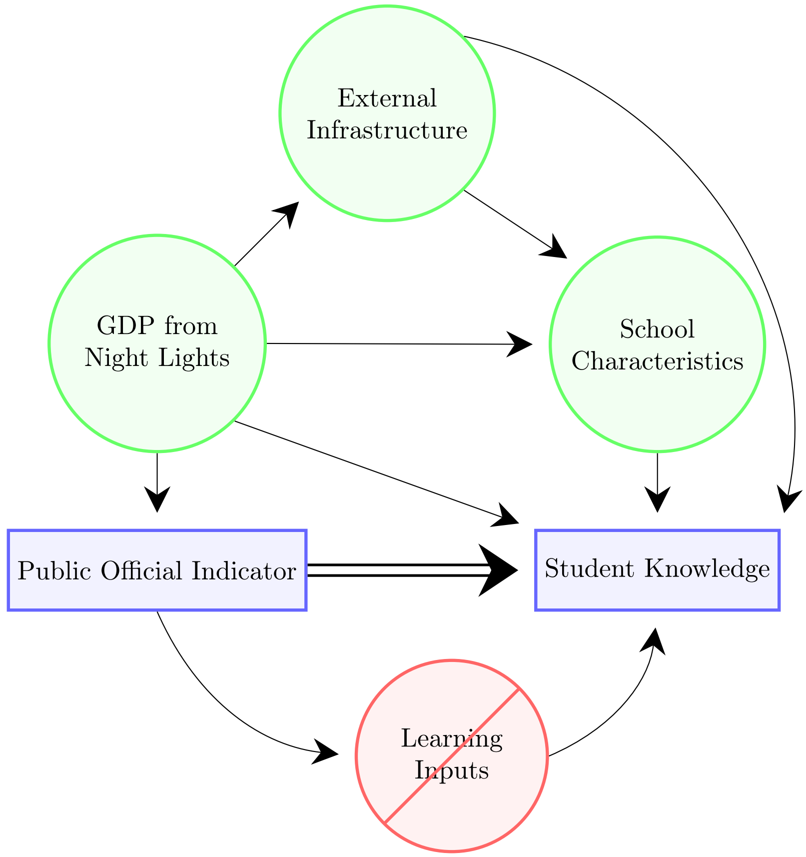

Figure 1: DAG

4.1.1 Ordinary Least Squares

We begin our estimation strategy with a simple baseline Ordinary Least Squared (OLS) model with controls in the form of:

$$Y_i = \beta_0 + \beta_1 D_{ji} + \beta_2 \mathbf{X}_i + \epsilon_i \tag{1}$$

Where Y_i is the learning outcome of school i. D_{ji} is the score of Public Official Indicator j for school i. The vector of covariates is given by $\mathbf{X}_i$. The OLS framework is naturally weak for analysis of this nature due to its sensitivity to outliers. OLS assumes that the errors in the data are normally distributed and have constant variance. Outliers or extreme observations can significantly impact the OLS estimates, leading to biased and inefficient results. Another issue with OLS is the endogeneity between the covariates and the error term. OLS assumes that the covariates are exogenous. However, in many empirical studies such as ours, endogeneity may exist, where the covariates are correlated with the error term, violating this OLS assumption.

4.1.2 Instrumental Variables

These controls are certainly not robust enough to claim a causal effect. Even with a large number of controls, the effect of indicators on learning outcomes is a very complex relationship. Omitted variable bias and unconfoundedness are still a concern. One approach to address this problem is through the use of an IV. This method involves identifying a variable correlated with our primary independent variable (public official index) but unconfounded with the error term. With a valid IV, we can find a local average effect that is less biased than our limited controls least squares estimate. So, in addition to ordinary least squares, we provide an IV approach as follows:

$$D_i = \delta_0 + \delta_1 \mathcal{Z}_i + \delta_2 \mathbf{X}_i + u_i \tag{2}$$

Equation 2 is the first stage of the two-stage least squares method. The variable notation remains the same; however, we estimate with public official indicators D_i as the dependent variable and include an IV $\mathcal{Z}_i$ along with our array of controls. We then take the fitted values from this estimation $\hat{D}_i$ and use them in place of the original values for public officials indicators as shown in equation 3 below:

$$y_i = \alpha_0 + \alpha_1 \hat{D}_i + \alpha_2 \mathbf{X}_i + \epsilon_i \tag{3}$$

An IV approach is a natural remedy for bias and confounding factors (Angrist and Krueger, 2001). A valid instrument must satisfy three criteria to provide an unbiased estimator. First, the instrument must have a causal effect on the endogenous explanatory variable of choice. Second, the instrument must affect the outcome variable only through the endogenous explanatory variable of choice, often called the "exclusion restriction." Third, the instrument is not a confounder; simply put, the instrument is uncorrelated with the dependent variable and the error term.

The drawback of an IV approach is that results will be biased when the instrument is not sufficiently strong or if the instrument is invalid. Should an instrument be considered weak, it would not provide sufficient variation to the input variable of interest, leading to a biased estimate. While the correlation between the instrument and the endogenous explanatory variable can have tests to determine if the instrument is sufficiently strong, we cannot directly test the exclusion restriction. Violation of the exclusion restriction is the most common reason behind an invalid IV estimation. (Angrist and Krueger, 1995; Hahn and Hausman, 2003). Thus, the researcher must possess a strong understanding of the system under investigation.

We utilize the distance between a school and its local administrative office as an instrument. We argue that this distance only affects learning outcomes through the bureaucratic efficiency of the district. More remote schools are more isolated from the administration and require a more efficient district office to receive the resources needed to fuel learning. Conversely, schools geographically closer to the district office are more likely to receive more attention and resources. Geography can be thought of as randomly assigned, so it should not affect the school's characteristic covariates. Thus, the distance is a candidate for a valid instrument.

IV approaches have been applied extensively in the field of education research. Sabarwal et al. (2014) employ an IV estimation approach to show that a textbook distribution program had little to no impact on student learning outcomes or attendance. An IV approach is used to identify the impacts of education within different demographic groups (Cannonier and Burke, 2022). Wodon and Ying (2009) used an instrument to assess the quality of public versus private schooling. However, while several papers have used IV approaches to study academic learning outcomes, there have not been any studies applying an IV technique to administration quality to identify impacts on educational outcomes. Additionally, many studies have used survey data in a similar context to our analysis to assess administration quality and its subsequent impacts on learning (Hallak and Poisson, 2005; Heck et al., 2000; Waugh, 2002).

4.1.3 Double Machine Learning

When all potential confounders are observed within a given data set, two estimation problems arise. First, the set of confounders is high-dimensional (i.e., there are more covariates than observations), which makes classical statistical approaches invalid. Second, the effects of the covariates on the treatment and outcome cannot be satisfactorily modeled by a parametric function. Both cases are addressed via DML (Chernozhukov et al., 2018). The DML procedure alleviates these issues mentioned above by first predicting the outcome of interest from the covariates, then predicting the treatment of interest from the covariates. Once this is done, the DML method combines these two predictive models in a final stage estimation which models the heterogeneous treatment effect. The strength of DML lies in its flexibility; that is to say, various combinations of traditional Machine Learning methods (e.g., L1/L2, SVM, Random Forest, etc.) can be used to make the two predictions mentioned above while removing any regularization bias that is commonly associated with such techniques.

We apply the DML procedure on the following Partial Linear Regression (PLR) as in Robinson (1988):

$$Y_i = D_{ji}\theta_0 + g_0(\mathbf{X}_i) + U_i, \qquad E[U \mid X, D] = 0 \tag{4}$$

$$D_{ji} = m_0(\mathbf{X}_i) + V_i, \qquad E[U \mid X] = 0 \tag{5}$$

Y_i is the learning outcome of school i, D_{ji} is the aggregated score of the Public Official Indicator j for school i. $\mathbf{X}_i$ is the vector of all covariates (called nuisance parameters in the context of DML). One benefit of the DML approach is that it allows for arbitrary single machine learning algorithms to be used for the two predictive tasks, namely m_0 and g_0. Another benefit of the DML framework is that it allows for the inclusion of higher-order covariates and interactions between covariates. We use the `PolynomialFeatures` class from `sklearn` to generate 3rd-order polynomials and interactions of all covariates for 3,276 nuisance parameters. The intermediate prediction models in our analysis utilize linear models fitted by minimizing the regularized empirical loss with Stochastic Gradient Descent (SGD). SGD is preferred to standard gradient descent as it is computationally less intensive and allows large sets of nuisance parameters to be computed more efficiently within the DML algorithm. However, the size of our dataset limits us to only creating two folds for the intermediate prediction models (i.e., a training set and a test set). Moreover, to counteract the size limitation of our data, we utilize the `DoubleML` build-in bootstrapping method to generate robust estimates. We use a standard bootstrap that iterates through the DML algorithm 500,000 times.

The DML framework for causal inference was introduced by Chernozhukov et al. (2018), and rapid theoretical development has been made since (Chang, 2020; Chernozhukov et al., 2017, 2018; Mackey et al., 2018). However, few papers utilize the DML framework to produce empirical results, such as (Knaus, 2022). To our knowledge, only Felderer et al. (2023) applies the DML framework to survey data. Moreover, since the introduction of the DML framework, only McNamara (2020) utilizes DML in the context of educational outcomes. The author estimates short-term individual returns to higher education in the United Kingdom. Therefore, our report is the first of its kind that explores the use of DML in the combined context of survey data and primary school learning outcomes.

4.2 Index Creation and Aggregation

Our data is qualitative by nature, so judgment calls must be made when it comes to index compilation. The World Bank indices are compiled by taking the averages of survey questions. Public officials are geographically linked to schools by district and are then averaged. These two processes create two opportunities for index adjustment. We present two adjustments to exemplify an index creator's possible choices.

Firstly, when linking public officials to schools and aggregating, a weighted mean can give certain officials more of a voice. For our purposes, we will weigh the aggregation by a score each enumerator gave the respondent at the end of each interview. The enumerator answered the question, "Did the respondent appear knowledgeable about the work environment and their organization as a whole?" on a scale from 1 to 3. In this case, 1 is less overall knowledge, and 3 is expert knowledge. The thought process here is that while all public official responses are important, more knowledgeable officials have a better understanding of bureaucratic operations as a whole.

Additionally, when compiling questions into indicators and sub-indicators, the question responses can be weighted using an alternative to the standard arithmetic mean. We follow the framework of quasi-arithmetic means used by Ferrant et al. (2020) that prevents very damaging high discrimination scores in some variables from being negated by very low scores in others.

$$M_q(x) = f^{-1}\left(\frac{f(x_1) + ... + f(x_n)}{n}\right) \tag{6}$$

Equation 6 shows how quasi-arithmetic means are calculated. The function f can be any continuous and injective function. Similarly to Ferrant et al. (2020), we implement a function that will prevent low-scoring public official questions from being negated by other high-scoring questions. The intuition is that lower scores damage overall efficiency more than high scores help. We choose the inverse ($\frac{1}{x}$) function for the quasi-arithmetic mean to achieve this.

Our results will include four differently calculated indices. First, in Panel A, we include one in which the public official indicators are compiled with the arithmetic mean of relevant questions and a simple average of public officials for aggregation. This will serve as the default method—. This is how the World Bank initially aggregated and compiled the indices. Second, we maintain the arithmetic mean compilation in Panel B but apply a higher weight for "more knowledgeable" officials during aggregation. Third, in Panel C, we apply the quasi-arithmetic mean compilation method to punish low scores as described above. Finally, in Panel D, we simultaneously apply both the weighted aggregation method and the quasi-arithmetic mean compilation method.

Table 2: Summary Stats for Indices (N = 232)

Panel A: Arithmetic Mean with Standard Aggregation	Mean	Std. Dev.	Min	Max
Bureaucratic Efficiency	3.66	0.25	3.17	4.00
Impartial Decision Making	3.74	0.32	2.61	4.08
Quality of Bureaucracy	3.95	0.24	3.22	4.54
Mandates and Accountability	3.71	0.49	2.62	4.26
National Learning Goals	3.23	0.42	2.40	3.79
Panel B: Arithmetic Mean with Weighted Aggregation				
Bureaucratic Efficiency	3.72	0.21	3.35	4.02
Impartial Decision Making	3.78	0.31	2.78	4.13
Quality of Bureaucracy	3.97	0.21	3.47	4.54
Mandates and Accountability	3.75	0.46	2.74	4.26
National Learning Goals	3.25	0.39	2.55	3.77
Panel C: Quasi-Arithmetic Mean with Standard Aggregation				
Bureaucratic Efficiency	2.84	0.26	2.41	3.18
Impartial Decision Making	2.99	0.40	1.76	3.43
Quality of Bureaucracy	3.33	0.31	2.72	4.17
Mandates and Accountability	3.28	0.53	2.20	4.04
National Learning Goals	2.43	0.36	1.79	3.22
Panel D: Quasi-Arithmetic Mean with Weighted Aggregation				
Bureaucratic Efficiency	2.87	0.24	2.48	3.19
Impartial Decision Making	3.04	0.41	1.84	3.53
Quality of Bureaucracy	3.35	0.32	2.92	4.16
Mandates and Accountability	3.31	0.51	2.28	4.04
National Learning Goals	2.44	0.33	1.87	3.18

[1] We calculated these indices from a survey given to public officials with questions falling into four categories: national learning goals, mandates and accountability, impartial decision-making, and quality of bureaucracy. Bureaucratic Efficiency is an index that aggregates all questions from the four categories.[2]See the code documentation for explicit variable definitions.

Summary statistics for these indices are shown in the separate panel of Table 2. Comparing Panel A to Panel B shows that weighted aggregation based on perceived experience and knowledge leads to higher scores across the board, indicating that public officials deemed more knowledgeable are also more optimistic about bureaucratic efficiency. Comparing Panel A to Panel C shows the effect of applying the quasi-arithmetic mean described above. As expected, scores are lower across the board, resulting from the punishment placed upon low scores. Panel D combines both methods– as such, the scores are lower than Panel A due to the quasi-arithmetic mean, however not quite as low as Panel C because weighting based on knowledge raises the index scores.

5 Results

Our analysis begins with a naive OLS regression of official public indicators on student learning outcomes. Next, we build upon this model by including the 25 school characteristics shown in Table 1 as controls. We then explore an IV approach to further address any estimation bias. The final model we employ is a DML approach which provides the tightest confidence intervals and, thus, is the most precise analysis method. Our results demonstrate the validity of applying DML in an administrative assessment context. In the following discussion of the results, we first focus on Table 3, the standard aggregation method, and question compilation using arithmetic means.

Table 3: Index Scores on Student Assessment Scores

Indices	Simple OLS (1)	Controls OLS (2)	Distance IV (3)	DML PLR SGD (4)
Arithmetic Mean with Standard Aggregation				
BE	8.80**	4.54	127.22	0.58***
	(1.39 - 16.22)	(-4.22 - 13.29)	(-257.27 - 511.70)	(0.54 - 0.62)
IDM	2.18	3.44	25.42	1.21***
	(-2.49 - 6.85)	(-2.76 - 9.64)	(-10.57 - 61.42)	(1.01 - 1.41)
QOB	8.17**	8.02*	55.00	-0.22***
	(1.23 - 15.10)	(-0.24 - 16.28)	(-37.83 - 147.84)	(-0.26 - -0.19)
MA	4.39**	1.72	-184.18	0.54***
	(0.42 - 8.36)	(-2.99 - 6.44)	(-1524.18 - 1155.82)	(0.50 - 0.58)
NLG	2.10	-0.37	-48.45	0.47***
	(-2.24 - 6.43)	(-5.63 - 4.90)	(-141.87 - 44.98)	(0.32 - 0.61)
Observations	297	232	232	232

95% Confidence Intervals are in parenthesis.
*** p<0.01, ** p<0.05, * p<0.1
BE refers to the Bureaucratic Efficiency indicator, IDM refers to Impartial Decision Making, QOB refers to Quality of Bureaucracy, MA refers to Mandates and Accountability, NLG refers to National Learning Goals.

Column 1 presents the results of the naive OLS regression—with confidence intervals reported in parenthesis. This estimates that a 1 ranking increase in our bureaucratic efficiency indicator is associated with an 8.80 percentage point increase in student knowledge assessment scores. The difference between the minimum and maximum school district scores is 0.8, which estimates a 7 percentile point gap between high and low-scoring districts. The precision of this estimate is a concern: the 95% confidence interval spans from 1.39 to 16.22 percentage points, which serves as a baseline to examine the impact of adding our controls, implementing an IV, and DML. Column 2 of Table 3 builds upon the naive model, including the controls outlined in the data section, Table 1. Including these controls leads to a point estimate that is smaller in magnitude, a 4.54 percentage point increase. However, these controls may not be robust, and precision is not improved.

These results are still subject to confounders that affect both student learning outcomes and the quality of bureaucracy. Socioeconomic factors such as a high poverty rate negatively affect the learning outcomes of students and the quality of bureaucracy. Another possible confounder is income inequality between various districts. District offices in impoverished areas may lack the resources to equip all schools under their jurisdiction properly. Even with The Education Act of 2004 that made school free and compulsory (Wang, 2007), impoverished families may not have the financial resources to equip their children for educational success.

We next implement the IV approach because of these concerns about latent confounders. A valid instrument variable allows the identification of the causal relationship between public official indicators and assessment scores depicted in Figure 1. In particular, there is a backdoor into student knowledge through the aforementioned unobserved characteristics. In column 3 of Table 3, we present the results of expanding our model using our instrument. The instrument we use is the distance from the school to the local administrative district office. However, when applying the IV, the point estimates become economically unrealistic—. An increase of 1 in our bureaucratic efficiency ranking is associated with a 127 percentage point increase in student assessment scores. This unrealistic point estimate is due to a massive loss of precision: the 95% confidence interval spans from -257.27 to 511.70. These large confidence values persist for all sub-indicators as well. The first-stage regression results are in Appendix Table 5. The distance instrument shows a very weak effect on public official indicators, and the overall significance of the model is relatively low for all indicators. Additionally, these results are imprecise because distance likely violates the unconfoundedness assumption and, therefore, may not be randomly assigned. The distance to administration is also very likely to affect other school characteristics included in the regression as well as unobserved confounders.

As the IV regression results proved problematic, we utilized a DML approach to yield more precise and reliable causal estimates. As seen in Column 4, a 1 ranking increase in our BE indicator is associated with a 0.58 percentage point increase in student assessment scores. For all indicators, the confidence intervals exhibit noticeable precision improvements. These confidence intervals reinforce DML as a viable method when estimating the causal effect of bureaucracy on student assessments.

We have several concerns with the validity of our findings. While DML is the most reliable of the four methods tested, the method is not a bulletproof framework. Like all econometric models, if the set of covariates is considered to be bad controls, as discussed by Cinelli et al. (2020), the DML method will also produce biased results. Similarly, the DML estimation will be unreliable if the intermediate models used in the DML procedure are ill-equipped to predict either learning outcomes or our indicators. Therefore caution must be exercised when constructing the DML framework. There are also concerns centered around the size of the data. Due to public officials serving all schools within their respective districts, the index aggregation process assigns each school within a district the same public official scores, which reduces the variation in our public official indicators from 232 down to 14 —the number of districts sampled. This lack of variation concerning our indicators inhibits our estimator and obfuscates any variation within each district.

5.1 Alternative Indices

This section makes the following assumptions: First, we assume that more tenured public officials are more knowledgeable in evaluating bureaucracy. Second, one cannot assume that the symmetrical property of the arithmetic mean holds. In other words, low scores in certain evaluation metrics could have a greater detraction effect on bureaucratic efficiency than the positive effect associated with higher scores in other metrics. It is important to emphasize that these questions and indices attempt to measure qualitative data using quantitative measures. As such, discretion from the survey creator and index compiler is required. There is no method to prove which index most closely matches the underlying truth of the world.

Panel A-C of Table 4 incorporates these assumptions both individually and jointly. This serves as a robustness check for the results of our proposed models. The pattern of model results remains consistent with the default index in Table 3: the accuracy shown by the confidence intervals of both OLS models remains wide in all three alternative indices. Similarly, the estimates for the IV models are unrealistically large due to extreme variances and confidence intervals. Our DML procedure continues to remain the most precise estimator (smallest confidence intervals) throughout all alternative index panels and sub-indicators.

Despite the precision of these results, it is important to note that the economic significance of these estimates is questionable. Table 2 shows that our index scores range from minimum to maximum and only span 1 to 2 ranks. Thus, our models predict that the schools with the lowest index scores are within 1 to 2 percentage points on assessment scores compared to the highest-ranking schools, which is also a consequence of aggregating public officials by districts which limits the variation in our indices, as mentioned above. However, our quasi-arithmetic method using the inverse function does result in negative estimates for our bureaucratic efficiency estimates, which is contrary to what is expected given the framework of this index created by the World Bank (WB, 2017b) and may indicate that the punishment of low scores is not desirable.

Table 4: Alternative Index Scores on Student Assessment Scores

Indices	Basic OLS (1)	Controls OLS (2)	Distance IV (3)	DML PLR SGD (4)
Panel A: Arithmetic Mean with Weighted Aggregation				
BE	9.89**	6.75	153.87	0.96***
	(1.58 - 18.21)	(-3.07 - 16.57)	(-335.07 - 642.82)	(0.86 - 1.07)
IDM	1.43	3.37	28.97	0.53***
	(-3.50 - 6.36)	(-3.12 - 9.85)	(-13.16 - 71.10)	(0.42 - 0.63)
QOB	6.57	8.29*	91.49	-0.25***
	(-1.64 - 14.82)	(-1.32 - 17.90)	(-100.33 - 283.31)	(-0.32 - -0.19)
MA	5.58***	2.68	-259.46	-0.95***
	(1.41 - 9.75)	(-2.28 - 7.64)	(-2799.47 - 2280.54)	(-1.02 - -0.88)
NLG	2.76	0.63	-50.84	1.27***
	(-1,84 - 7.37)	(-4.94 - 6.19)	(-149.36 - 47.68)	(0.97 - 1.57)
Panel B: Quasi-Arithmetic Mean with Standard Aggregation				
BE	0.86	-0.24	-153.06	-0.24***
	(-6.23 - 7.95)	(-8.34 - 7.86)	(-705.01 - 398.90)	(-0.27 - -0.20)
IDM	-0.21	0.54	20.66	0.42***
	(-4.08 - 3.66)	(-4.49 - 5.57)	(-8.75 - 50.08)	(0.38 - 0.45)
QOB	3.30	0.98	-99.93	0.01**
	(-2.71 - 9.30)	(-6.08 - 8.04)	(-403.53 - 203.68)	(0.006 - 0.013)
MA	3.18*	2.17	-113.42	0.14***
	(-0.45 - 6.82)	(-2.10 - 6.43)	(-648.97 - 422.14)	(0.13 - 0.16)
NLG	-1.21	-5.49*	-17.87	0.51***
	(-6.09 - 3.67)	(-11.19 - 0.21)	(-40.97 - 5.22)	(0.37 - 0.65)
Panel C: Quasi-Arithmetic Mean with Weighted Aggregation				
BE	1.11	0.66	-90.31	-0.13***
	(-6.49 - 8.69)	(-7.91 - 9.22)	(-292.45 - 111.82)	(-0.17 - -0.09)
IDM	-0.54	0.95	20.85	0.75***
	(-4.35 - 3.27)	(-3.97 - 5.87)	(-8.96 - 50.66)	(0.64 - 0.85)
QOB	0.72	-0.06	-54.76	-0.87***
	(-5.55 - 6.99)	(-7.09 - 6.96)	(-159.97 - 50.44)	(-0.97 - -0.76)
MA	4.68**	3.23	-102.63	-0.38***
	(0.92 - 8.43)	(-1,22 - 7.68)	(-528.23 - 322.96)	(-0.43 - -0.33)
NLG	-0.91	-5.24*	-17.65	-0.12
	(-6.06 - 4.25)	(-11.26 - 0.79)	(-40.40 - 5.09)	(-0.38 - 0.14)
Observations	297	232	232	232

95% Confidence Intervals are in parenthesis.
*** p<0.01, ** p<0.05, * p<0.1
BE refers to the Bureaucratic Efficiency indicator, IDM refers to Impartial Decision Making, QOB refers to Quality of Bureaucracy, MA refers to Mandates and Accountability, NLG refers to National Learning Goals.

Index creation is one of many concerns. Another concern revolves around the nature of survey data. There is an ever-present worry of response and recall bias- there is no way to guarantee honesty and correct efficiency evaluations from public officials. The enumeration process alleviates a portion of this concern: enumerators decide how to score each category after a series of questions designed to elicit the state of operations. However, this does not solve the problem entirely and furthermore introduces enumeration bias.

The scope of our results is limited to Sierra Leone. The country has a unique combination of government, society, and history compared to other countries in the region. Moreover, the unique combination of Sierra Leone's one-party government prior to 1991, the extensive use of child soldiers over the course of a 10-year Civil War, and the vast rebuilding of post-war infrastructure limit our findings to Sierra Leone only. Therefore, while our study provides important insights into the effects of these public official indicators on student knowledge in Sierra Leone, caution should be exercised in generalizing any findings to broader populations and settings. Nevertheless, the framework in this report provides a method for establishing links between public bureaucracy and educational outcomes using survey data, regardless of country.

6 Conclusion

This report examines the impact of educational bureaucracy on learning outcomes by analyzing a survey of public officials in Sierra Leone conducted by the World Bank. The survey captures the opinions of public officials regarding administration. It is comprised of 52 questions separated into four categories: National Learning Goals (NLG), Mandates and Accountability (MA), Quality of Bureaucracy (QOB), and Impartial Decision-Making (IDM). We compile all questions into an overall Bureaucratic Efficiency (BE) indicator. Factors such as sample size and unobserved confounders make a reliable estimation of the link between bureaucracy and learning outcomes using Ordinary Least Squares (OLS) difficult to achieve. An instrumental variable (IV) approach is a classic remedy to these issues. However, the lack of a clear and valid instrument limits the precision of IV results. We then use a Double Machine Learning (DML) approach due to our ill-conditioned data—large number of controls or covariates and small sample size. To ensure the validity of our findings, we conduct various robustness checks by employing alternative mean aggregation and index creation techniques.

Our estimation method reinforces the validity of applying DML models to estimate learning outcomes precisely. DML produces precise confidence intervals, starkly contrasting those produced by OLS or IV models. Overall, our findings suggest that using DML to estimate effects using this kind of survey data is a valid approach for more precision.

There are caveats to our findings. First, the sample size of individuals interviewed limits the reliability of our findings; a larger sample of public officials and student assessment scores would provide more transparency as to the validity of our findings. Second, the data-generating process is missing key attributes regarding the student's socioeconomic status. Obtaining socioeconomic data will enable future model iterations to control for key elements of a student's background, leading to tighter estimates. Finally, inherent assumptions are required to create survey questions to measure qualitative data and compile indices from them. The caveats mentioned here should be considered in future survey iterations.

Given our novel findings, future research should consider expanding on the current survey methods for public officials and school officials. By doing so, relationships such as connecting bureaucratic efficiency to teacher and student absenteeism can be established. Similarly, future versions of this survey should strive to include multiple cross-sections to allow for analysis of public official indicators to vary with time. Similarly, future analysis could build upon these results by including multiple countries. However, researchers should be cautious about country-specific variation. Future research should take these considerations into account to illuminate potential policy options to increase the effectiveness of education administration.

References

Akhtari, M., D. Moreira, and L. Trucco (2022, February). Political turnover, bureaucratic turnover, and the quality of public services. *American Economic Review 112*(2), 442–93.

Amowitz, L. L., C. Reis, K. H. Lyons, B. Vann, B. Mansaray, A. M. Akinsulure-Smith, L. Taylor, and V. Iacopino (2002). Prevalence of war-related sexual violence and other human rights abuses among internally displaced persons in sierra leone.

Angrist, J. D. and A. B. Krueger (1995). Split-sample instrumental variables estimates of the return to schooling. *Journal of Business & Economic Statistics 13*(2), 225–235.

Angrist, J. D. and A. B. Krueger (2001, December). Instrumental variables and the search for identification: From supply and demand to natural experiments. *Journal of Economic Perspectives 15*(4), 69–85.

Asongu, S. A. and N. M. Odhiambo (2019). Enhancing ict for quality education in sub-saharan africa. *Education and information technologies 24*, 2823–2839.

Banerjee, A. V., R. Banerji, E. Duflo, R. Glennerster, and S. Khemani (2010). Pitfalls of participatory programs: Evidence from a randomized evaluation in education in india. *American Economic Journal: Economic Policy 2*(1), 1–30.

Banerjee, A. V., S. Cole, E. Duflo, and L. Linden (2007). Remedying education: Evidence from two randomized experiments in india. *The Quarterly Journal of Economics 122*(3), 1235–1264.

Banya, K. (1991). Economic decline and the education system: the case of sierra leone.

Banya, K. (1993). Illiteracy, colonial legacy and education: the case of modern sierra leone.

Barankay, I. and B. Lockwood (2007). Decentralization and the productive efficiency of government: Evidence from swiss cantons. *Journal of public economics 91*(5-6), 1197–1218.

Bau, N. and J. Das (2017). *The Misallocation of Pay and Productivity in the Public Sector: Evidence from the Labor Market for Teachers.* World Bank.

Beaman, L., E. Duflo, R. Pande, and P. Topalova (2012). Female leadership raises aspirations and educational attainment for girls: A policy experiment in india. *Science 335*(6068), 582–586.

Betancourt, T. S., S. Simmons, I. Borisova, S. E. Brewer, U. Iweala, and M. De La Soudière (2008). High hopes, grim reality: Reintegration and the education of former child soldiers in sierra leone.

Borcan, O., M. Lindahl, and A. Mitrut (2017). Fighting corruption in education: What works and who benefits? *American Economic Journal: Economic Policy 9*(1), 180–209.

Callen, M., S. Gulzar, A. Hasanain, M. Y. Khan, and A. Rezaee (2023). The political economy of public sector absence. *Journal of Public Economics 218*, 104787.

Cannonier, C. and M. G. Burke (2022). The impact of education on household decision-making among women in sierra leone. In *The Palgrave Handbook of Africa's Economic Sectors*, pp. 217–254. Springer.

Chang, N.-C. (2020). Double/debiased machine learning for difference-in-differences models. *The Econometrics Journal 23*(2), 177–191.

Chernozhukov, V., D. Chetverikov, M. Demirer, E. Duflo, C. Hansen, and W. Newey (2017, May). Double/debiased/neyman machine learning of treatment effects. *American Economic Review 107*(5), 261–65.

Chernozhukov, V., D. Chetverikov, M. Demirer, E. Duflo, C. Hansen, W. Newey, and J. Robins (2018, 01). Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal 21*(1), C1–C68.

Chernozhukov, V., W. K. Newey, and J. Robins (2018). Double/de-biased machine learning using regularized riesz representers.

Cinelli, C., A. Forney, and J. Pearl (2020). A crash course in good and bad controls.

Clark, D. (2009). The performance and competitive effects of school autonomy. *Journal of political Economy 117*(4), 745–783.

Duflo, E., P. Dupas, and M. Kremer (2015a, September). Education, hiv, and early fertility: Experimental evidence from kenya. *American Economic Review 105*(9), 2757–97.

Duflo, E., P. Dupas, and M. Kremer (2015b). School governance, teacher incentives, and pupil–teacher ratios: Experimental evidence from kenyan primary schools. *Journal of public Economics 123*, 92–110.

Duflo, E., P. Dupas, and M. Kremer (2021). The impact of free secondary education: Experimental evidence from ghana. Technical report, National Bureau of Economic Research.

Faguet, J.-P. and F. Sánchez (2014). Decentralization and access to social services in colombia. *Public choice 160*, 227–249.

Felderer, B., J. Kueck, and M. Spindler (2023). Using double machine learning to understand nonresponse in the recruitment of a mixed-mode online panel. *Social Science Computer Review 41*(2), 461–481.

Ferrant, G., L. Fuiret, and E. Zambrano (2020). The social institutions and gender index (sigi) 2019: A revised framework for better advocacy.

Galiani, S., P. Gertler, and E. Schargrodsky (2008). School decentralization: Helping the good get better, but leaving the poor behind. *Journal of public economics 92*(10-11), 2106–2120.

Glennerster, R., S. Imran, and K. Whiteside (2006). Baseline report on the quality of primary education in sierra leone: Teacher absence, school inputs, and school supervision in district education committee and government-assisted schools in sierra leone.

Goldstein, M. and E. Dabla-Norris (2004). The distributional bias of public education: Causes and consequences.

Hahn, J. and J. Hausman (2003, May). Weak instruments: Diagnosis and cures in empirical econometrics. *American Economic Review 93*(2), 118–125.

Hallak, J. and M. Poisson (2005). Ethics and corruption in education: an overview. *Journal of education for international development 1*(1), 1–3.

Haller, E. J. et al. (1994). Does graduate training in educational administration improve america's schools? another look at some national data.

Hallinger, P. and R. H. Heck (1996). Reassessing the principal's role in school effectiveness: A review of empirical research, 1980-1995. *Educational administration quarterly 32*(1), 5–44.

Hallinger, P. and R. H. Heck (2011). *Leadership and student learning outcomes*. Sage.

Heck, R. H., L. K. Johnsrud, and V. J. Rosser (2000). Administrative effectiveness in higher education: Improving assessment procedures. *Research in higher education 41*, 663–684.

Heywood, P. M. (2014). Measuring corruption: perspectives, critiques and limits. In *Routledge handbook of political corruption*, pp. 136–153. Routledge.

Hill, S. and T. Chalaux (2011). Improving access and quality in the indian education system.

Hinton, S. (2009). Questions in the answers to primary school educational reconstruction in sierra leone.

Jacob, O. N., H. F. Josiah, and A. T. Solomon (2021). Effects of corruption on public universities administration in nigeria. *Emergent: Journal of Educational Discoveries and Lifelong Learning 2*(7), 1–14.

Jang, S. Y. (2012). The causes of the sierra leone civil war.

Keen, D. (2003). Greedy elites, dwindling resources, alienated youths the anatomy of protracted violence in sierra leone.

Knaus, M. C. (2022). Double machine learning-based programme evaluation under unconfoundedness. *The Econometrics Journal 25*(3), 602–627.

Kremer, M., C. Brannen, and R. Glennerster (2013). The challenge of education and learning in the developing world. *Science 340*(6130), 297–300.

Lai, F., R. Luo, L. Zhang, X. Huang, and S. Rozelle (2015). Does computer-assisted learning improve learning outcomes? evidence from a randomized experiment in migrant schools in beijing. *Economics of Education Review 47*, 34–48.

Leithwood, K., A. Harris, and D. Hopkins (2008). Seven strong claims about successful school leadership. *School leadership and management 28*(1), 27–42.

Leone, R. S. (2001). Interim poverty reduction strategy paper.

Mackey, L., V. Syrgkanis, and I. Zadik (2018, 10-15 Jul). Orthogonal machine learning: Power and limitations.

Maclure, R. and M. Denov (2009). Reconstruction versus transformation: Post-war education and the struggle for gender equity in sierra leone.

McDermott, P. and N. Allen (2015). Successes and challenges of implementing a teacher education project in rural sierra leone.

McKay, S. (2005). Girls as "weapons of terror" in northern uganda and sierra leonean rebel fighting forces.

McNamara, S. (2020). Returns to higher education and dropouts: A double machine learning approach.

Miguel, E. and M. Kremer (2004). Worms: identifying impacts on education and health in the presence of treatment externalities. *Econometrica 72*(1), 159–217.

Muralidharan, K. and V. Sundararaman (2013). Contract teachers: Experimental evidence from india. Technical report, National Bureau of Economic Research.

Murtin, F. (2013). Improving education quality in south africa.

Olken, B. A. and R. Pande (2012). Corruption in developing countries. *Annu. Rev. Econ. 4*(1), 479–509.

Pearl, J. (2009). *Causality.* Cambridge university press.

Peters, K. and P. Richards (1998). 'why we fight': Voices of youth combatants in sierra leone.

Reno, W. (2003). Political networks in a failing stage the roots and future of violent conflict in sierra leone.

Riddell, W. C. and X. Song (2011). The impact of education on unemployment incidence and re-employment success: Evidence from the us labour market. *Labour Economics 18*(4), 453–463.

Robinson, P. M. (1988). Root-n-consistent semiparametric regression.

Robinson, V. M., M. Hohepa, and C. Lloyd (2007). *School leadership and student outcomes: Identifying what works and why*, Volume 41. Australian Council for Educational Leaders Winmalee.

Sabarwal, S., D. Evans, and A. Marshak (2014). The permanent input hypothesis: the case of textbooks and (no) student learning in sierra leone.

Smidova, Z. (2019). Educational outcomes: A literature review of policy drivers from a macroeconomic perspective.

Taylor, L. (2003). " we'll kill you if you cry": Sexual violence in the sierra leone conflict.

UN (2000). United nations millennium development goals.

UN (2015). The 2030 agenda for sustainable development.

Wang, L. (2007). Education in sierra leone: Present challenges, future opportunities.

Waugh, R. F. (2002). Academic staff perceptions of administrative quality at universities. *Journal of Educational Administration 40*(2), 172–188.

WB (2017a). World development report 2017: Governance and the law.

WB (2017b). World development report 2018: Learning to realize education's promise.

Wodon, Q. and Y. Ying (2009). Literacy and numeracy in faith-based and government schools in sierra leone.

A Appendix

A.1 Supplemental IV Table

Table 5: First-Stage for Default Indices

VARIABLES	(1) BE	(2) IDM	(3) QOB	(4) MA	(5) NLG
Distance	-0.00	-0.00**	-0.00	0.00	0.00
Absence Rate	-0.00	-0.00*	-0.00*	-0.00	0.00
Student Attendance	-0.00	0.00	-0.00	-0.00	0.00
Students Enrolled	-0.00	0.00***	-0.00	-0.00	-0.00*
Content Proficiency	0.00*	0.00	0.00	0.00**	0.00
ECD Student Proficiency	0.00	-0.00	0.00	0.00	0.00
Infrastructure	0.00	-0.06***	-0.00	0.02	0.06**
Teach Score	-0.02	0.03	-0.03	-0.02	-0.05
Operational Management	0.01	0.00	0.04	-0.01	0.00
Instructional Leadership	-0.04**	0.01	-0.02	-0.08**	-0.07***
Principal Knowledge Score	0.05***	0.01	0.02	0.09***	0.07**
Principal Management	0.01	0.01	0.03	-0.02	0.02
Teacher Attraction	0.02	0.01	-0.01	0.01	0.07
Teacher Selection Deployment	0.01	-0.03	0.00	0.03	0.04
Teacher Support	0.03	-0.02	-0.02	0.12**	0.05
Teaching Evaluation	0.08***	0.02	-0.01	0.16***	0.14***
Teacher Monitoring	0.02	0.03	0.04	0.02	-0.01
Intrinsic Motivation	0.03	0.14***	0.06*	0.00	-0.09*
Standards Monitoring	0.02	-0.06**	-0.00	0.06	0.10***
School Monitoring	-0.02	-0.09***	-0.06**	0.05	0.02
School Management Clarity	-0.03	0.01	-0.01	-0.06**	-0.05*
School Management Attraction	-0.05***	-0.00	0.01	-0.12***	-0.09***
School Selection Deployment	-0.01	0.00	0.01	-0.05	-0.02
School Support	-0.00	-0.00	-0.00	0.00	-0.00
Principal Evaluation	0.02	-0.09***	-0.01	0.08	0.10**
light_GDP	-0.01	0.03***	-0.01	-0.03*	-0.04***
Observations	232	232	232	232	232
R-squared	0.25	0.32	0.18	0.32	0.34
F-Stat	3.929	3.524	1.810	4.774	6.246
Prob > F	1.38e-08	2.10e-07	0.0124	5.00e-11	0
Degree of Freedom	205	205	205	205	205

Robust standard errors
*** p<0.01, ** p<0.05, * p<0.1