EdTech Hub

Clear evidence, better decisions, more learning.

FAB INC.

WORKING PAPER

Learning from Experience

A post-Covid-19 data architecture for a resilient education data ecosystem in Sierra Leone

Date August 2021

Authors Fab Inc.

DOI 10.5281/zenodo.5498054

THE WORLD BANK

UKaid from the British people

#EdTechHub @GlobalEdTechHub edtechhub.org

About this document

Recommended citation	Fab Inc. (2021). *Learning from Experience: A post-Covid-19 data architecture for a resilient education data ecosystem in Sierra Leone.* Working Paper 34. EdTech Hub. https://doi.org/10.5281/zenodo.5498054. Available at https://docs.edtechhub.org/lib/JTKTA5X5. Available under Creative Commons Attribution 4.0 International, https://creativecommons.org/licenses/by/4.0/.
Licence	Creative Commons Attribution 4.0 International https://creativecommons.org/licenses/by/4.0/. You are free to share (copy and redistribute the material in any medium or format) and adapt (remix, transform, and build upon the material) for any purpose, even commercially. You must give appropriate credit, provide a link to the license, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.
Notes	EdTech Hub is supported by UK aid and the World Bank; however, the views expressed in this document do not necessarily reflect the views of the UK Government or the World Bank.
Acknowledgements	The Fab Inc. team wish to thank EdTech Hub for their support, particularly the in-country support. They also wish to thank the team at MBSSE, including the Directorate of Policy of Planning, for making the data available over the last few years, and engaging with the project.
Reviewers	Amy Bellinger, Joel Mitchell, Chris McBurnie, Katy Jordan

Abstract

This report learned from the experience of both Ebola and data management efforts to help inform post-Covid-19 recovery including a data architecture for a resilient education data ecosystem in Sierra Leone and elsewhere. This was built on the construction of a combined database containing Annual School Census data at the school level from 2015–2019, and at the district level from 2011–2013. Analysis of enrolment and Ebola case data highlighted the ongoing importance of geographical and economic factors in determining education access, over and above the short-term health impacts of such crises. This has also enabled a detailed consideration of the data architecture within Sierra Leone to inform ongoing stakeholder discussions in this area, as well as demonstrating the potential benefits of improved data structures for data analysis, visualisation, and use through a dashboard. This has been warmly received, with the government looking to take over ownership of the database, display the dashboard publicly, and inform ongoing policy development and implementation. Finally, key principles and step-by-step guides for replicating this work have been provided to support similar efforts in other countries which face comparable challenges to Sierra Leone in terms of both data availability and consistency, and the potential use of this data to support Covid-19 school reopening.

Contents

Abbreviations and acronyms

ASC	Annual School Census
BI	Business Intelligence
CSV	Comma Separated Values
DFID	Department for International Development
DSTI	Directorate of Science, Technology and Innovation
D4G	Data For Good
ELT	Extract, Load and Transform
EMIS	Education Management Information System
FCDO	Foreign, Commonwealth and Development Office
FQSE	Free Quality School Education
GIS	Geographic Information System
GoSL	Government of Sierra Leone
GPS	Global Positioning System
JSS	Junior Secondary School
OTPS	One Tablet Per School
MBSSE	Ministry of Basic and Senior Secondary Education
MICS	Multiple Indicator Cluster Surveys
ML	Machine Learning
MTHE	Ministry of Technical and Higher Education
RCT	Randomised Control Trial
SaaS	Software as a Service
SABER	Systems Approach for Better Education Results
SL	Sierra Leone
SQL	Structured Query Language
SSS	Senior Secondary School
TSC	Teaching Service Commission

Executive Summary

Sierra Leone is, unfortunately, no stranger to frequent and sometimes devastating shocks. Just six years ago, the country became embroiled in the battle against the deadly Ebola virus outbreak, with schools closed from July 2014 to April 2015. Sierra Leone's experience of rebuilding the education system after Ebola can help the country and the rest of the world with building back better following Covid-19.

In order to learn from past events, there needs to be reliable and comparable historical data to carry out longitudinal analysis. Unfortunately, in Sierra Leone, school census data — which is not only recorded in separate data files, but often in different formats year to year — was not previously developed to be used in this way. As a result, the research team partnered with the Ministry of Basic and Senior Secondary Education (MBSSE) to convert historically fragmented Annual School Census (ASC) datasets into a single fully integrated database, with two aims in mind:

a) To track the main trends in education enrolment, showing how the impact of the Ebola virus outbreak might guide school reopening following the current Covid-19 crisis and potential future crises

b) To inform, demonstrate and provide recommendations on how to use and manage an education system's data architecture

In response to the first goal, the integration of the ASC into a single data system enabled analysis of the enrolment changes and Ebola cases across the country around the school closures of the Ebola epidemic. This resulted in a number of findings. Firstly, the total enrolment numbers increased across the country between 2015 and 2019, but this picture is clouded by an increased coverage of schools identified in the 2015 ASC. Secondly, there was significant variation across districts with some districts reporting significantly fewer students in 2015, particularly at the primary level. Interestingly, there is no clear pattern between these changes and the Ebola cases per district. Thirdly, there was a greater adverse effect on enrolment for boys than for girls across each level of education. Nevertheless, localised issues and particular challenges for girls during school closures, such as risks of early pregnancy, should continue to be mitigated against. Fourthly, the potential of using data to predict vulnerability was demonstrated during the post-Ebola context to highlight techniques to support Covid-19 recovery.

Overall, the analysis of enrolment trends highlighted the ongoing importance of geographical and economic factors in determining education access, over and above the short-term health impacts of such crises.

The second aim of the study focused on the data architecture of the education sector in Sierra Leone and highlighted relevant data management principles to inform planning and contribute to ongoing stakeholder discussions in this area. The benefits of improved data structures, such as those developed as part of the combined 2015–2019 school-level database, were demonstrated with a dashboard that pulled from this database and enables easy and powerful data analysis and visualisation.

This has been warmly received and is already being widely used by stakeholders. The team is working with the government to support them to take over ownership of the database and integrate the dashboard in their public Education Data Hub and into data-driven processes more widely. This will support the implementation of the Government's ambitious reform agenda and several policies including the Radical Inclusion Policy, School Approvals Policy Guidelines and the School Infrastructure and Catchment Area Planning Policy.

Finally, these findings are also likely to be relevant to many other countries which face similar challenges to Sierra Leone, in terms of both data availability and consistency, and Covid-19 school reopening. Key principles and step-by-step guides for replicating this work in other contexts have been provided to maximise the use of this research and support similar efforts in other contexts.

1. Introduction

1.1 Background to study

Sierra Leone has experienced significant shocks in recent years, which have impacted upon access to education, including the deadly Ebola virus outbreak six years ago. However, Sierra Leone's experience of rebuilding the education system after Ebola can help the country and the rest of the world with building back better after Covid-19. Studying the impact of the Ebola virus outbreak on the education system will not only help in building an understanding around how such crises impact education systems but will also help in understanding what works, from a policy and practice perspective, in overcoming the short- and long-term shocks of such crises.

To be able to learn from past events, reliable historical data must be available for longitudinal analysis. Unfortunately, each round of school census data in Sierra Leone has been considered as independent, and no successful attempt has been made to compare and match these datasets to examine patterns over time. The lack of reliable longitudinal data has not only prevented the Government of Sierra Leone (GoSL) and its partners from understanding the impact of Ebola and other crises on the system but also kept the policymakers from knowing which policies have been most effective in mitigating and overcoming the long-term impacts of extended school closures due to Ebola.

In this study, the research team partnered with the Ministry of Basic and Senior Secondary Education (MBSSE) to convert historically fragmented Annual School Census (ASC) datasets, which were only available in separate data files and often in different formats, into a single fully integrated database with an interactive visual dashboard.

This fully integrated database would enable the GoSL to conduct longitudinal analysis on educational trends. Furthermore, with the recent collection of geospatial coordinates (from 2018 onwards), all education stakeholders would be able to use this database to track changes over time at the individual school-level. Such a tool would tremendously increase the options for data analysis, such as examining equitable access to schooling. Finally, the interactive visual dashboard would help policymakers to continue using this database to carry out analysis that would help them improve education planning and programming.

In the process of building this fully integrated database, the project will generate learnings on how standalone datasets can be merged to develop a multi-year database. The project will provide Sierra Leone with a working education database as well as a blueprint of the optimal structure of an easy-to-manage, dynamic education data architecture. This document will not only help Sierra Leone to sustain the integrated database but may also be useful to other countries wanting to carry out a similar exercise.

1.2 Purpose / aims of this paper

This paper has two main aims. Firstly, it examines Sierra Leone school census data to track the main trends in education enrolment, showing how the impact of the Ebola virus outbreak might guide school reopening following the current Covid-19 and potential future crises. Secondly, it provides recommendations on how to use and manage an education system's data architecture, drawing on our experience building a single fully integrated education database for Sierra Leone to examine such trends using previously fragmented ASC data. These two aims were based on the construction of a combined database containing school-level ASC data from 2015 to 2019, and analysis of district-level ASC data from 2011 (with the exception of 2014 when Ebola caused school closures).[1]

1.3 Research questions

To address the first aim of the paper (A), the following primary and secondary research questions were developed:

Primary research question

- RQA1. What lessons can be drawn from school-level changes in enrolment following Ebola, which can be applied to school reopening following Covid-19?

Secondary research questions

- RQA2. How did enrolment change across the country? Are there any hotspots of incidence?
- RQA3. Are there any differences by gender in terms of re-enrolment following Ebola?
- RQA4. Is it possible to use these patterns, and household survey data, to predict likely areas of vulnerability going forward?

Having the ASC data over a longer time period can help in answering key questions about how access, quality, and equity in education have changed. Even though learning outcome data is not available, the access (gender disaggregated enrolment) and quality (as perceived by inputs such as

[1] Since completing this research study, the team has added the 2020 ASC data to the integrated database.

teachers, facilities, and infrastructure) data in the ASC is invaluable in allowing policymakers to look backwards to plan forwards, and to observe how their policies and choices have influenced the system. This paper seeks to unpack some of these trends, with a particular focus on the before- / after-Ebola period, as well as on district-level trends in enrolment for nearly the last decade.

To address the second aim of the paper (B), the following research questions were examined:

- RQB1. What is the optimum education data architecture, and what are the key initial steps to take towards building such an architecture, given Sierra Leone's context, prior experiences, and the field constraints?
- RQB2. How can we demonstrate how the Annual School Census (ASC), built with a well-planned new data architecture, can help the government in improving education planning by using existing ASC data?
- RQB3. What are the key components / datasets for this system — what exists now, what is planned for the future, and how can both be linked?

This paper seeks to show the potential of an integrated multi-year school database and associated dashboard in improving education planning by answering these questions. Finally, it describes the processes and lessons learned in creating this multi-year database[2] and identifies any lessons for the wider Sierra Leone education data architecture.

1.4 Context of the study

Despite being beset by several shocks in recent years including the negative shocks of the legacy of the civil war and the Ebola virus outbreak, the Sierra Leone education system has made significant recovery in the form of rapid expansion of primary education in the 2000s and the recent surge in both primary and secondary enrolments as part of the 'Free Quality School Education' (FQSE) initiative which started in 2018. The current Covid-19 pandemic has once again negatively shocked the country's education system by causing widespread closure of schools. However, Sierra Leone is in the unenviable position of being a system that has in recent years already faced school closures due to a public health crisis, the Ebola crisis of 2014–2015. But the lessons from the educational impacts of the Ebola crisis and the effects of

[2] The term *database* is used in this paper to describe the combination of separate individual pieces of information (datasets) and aim to turn disparate datasets into one functioning database.

those school closures are quite limited, thereby creating a unique opportunity to add to the evidence base on how countries rebuild from shocks.

There is also a strong demand from the MBSSE and key stakeholders to improve the use of data in education management, and to understand more about the education system and how it has changed. A wave of innovation in education management and policies has been brought in by Minister of Education David Moinina Sengeh, who is a real champion of data-driven decision-making. This has been combined with an increase in public spending in education following the FQSE initiative. Overall, this has opened an unprecedented window of opportunity to improve evidence-based decision-making in education.

This ambition is shared across the education system, and there is demand from the MBSSE and Teaching Service Commission (TSC) to ensure the new Teacher Management Information System can help the GoSL drive decision-making and gender-sensitive workforce planning, and be integrated into wider system planning. There is also demand from the Local Education Group partners to improve the fragmentation of the system and ensure more reliable and regular data is available and used in policy making and monitoring.

1.5 What this paper adds to the knowledge base

This is the first paper, to our knowledge, that attempts to look at the Ebola data to understand how public health emergencies and school closures have led to system-level disparities in re-enrolments prior to Covid-19. While small scale research samples exist, no systematic study of country-wide census data has been conducted.

Similarly, the project which this paper reports on represents the first time that a low-income country has had a functioning school-level database spanning several years, which enables policymakers and researchers to identify patterns and changes over time. It also documents clearly the practical steps and considerations needed to construct such a database, including suggestions for how this can work in other contexts.

The paper further highlights the processes required for and lessons learned from creating an inclusive education management information database with an accessible and interactive data dashboard. In addition, the paper contributes to the practical discussions that are underway on the wider education data architecture, including ensuring alignment of ASC with the One Tablet Per School (OTPS) programme that is currently being rolled out, the Education Data Hub, and the forthcoming Teacher Management Information System.

1.6 Implications for policy and practice

This project has a wide range of implications for policy and practice, and is already having an impact. First, it enables lessons from previous long-term school closures during the Ebola epidemic to inform future programming and responses to the ongoing Covid-19 pandemic; principally, the consideration that indirect economic factors may play a larger role in determining the consequences on education outcomes than the more direct health challenges. Secondly, the research on data architecture within the Sierra Leone education sector can help to inform the ongoing discussions between stakeholders on data alignment and structures moving forwards; principally, in highlighting some of the key rules and lessons which should be prioritised during implementation.

Moreover, the key to gaining buy-in and translating this work into policy and programming has been the curation of the data into a simple dashboard, which allows the information to be visualised easily and processed quickly without the need for data skills. This separation of the datasets, the curation of dashboards, and then the exploration of the data is good practice when developing business intelligence tools in many sectors, as it allows a wider range of users to access key messages.

The interactive dashboard has already been presented to the GoSL and key stakeholders and a number of uses and follow-ups have been identified which will be covered in more detail in the policy implications section. These include:

- Making the dashboard publicly available as a formal addition to the Education Data Hub (https://educationdatahub.dsti.gov.sl/) and transferring storage of the database for national ownership
- The Minister of MBSSE presenting these outputs nationally with the President at the upcoming launch of the 2020 Annual School Census report
- The Minister of MBSSE presenting these outputs internationally at the Global Education Summit
- Building out specific dashboards for key stakeholders to enable data-for-delivery and data-for-storytelling tailored to specific audiences such as the MBSSE Director of Planning and Policy (to monitor implementation of the School Infrastructure and Catchment Area

Planning Policy) and the Teaching Service Commission (to focus on teacher data, with demand from the Chair through to the district level).

As this is an add-on to the research project, there is not the space to explore fully the choices and use of these tools in this report, but this is an area that has attracted substantial interest in Sierra Leone and beyond and one which merits further exploration.

The project has also enabled a detailed consideration of the data architecture within Sierra Leone to inform ongoing stakeholder discussions in this area. The Government acknowledges that data is a key input to implementing its ambitious reform agenda and new policies including the Radical Inclusion Policy, School Approvals Policy Guidelines and the School Infrastructure and Catchment Area Planning Policy. Focusing on the political and technical challenges of the data architecture and effective data management will maximise the impact of each of these policies and better enable Sierra Leone to meet the challenges of Covid-19 and potential future shocks.

1.7 Structure

The paper starts with a review of the literature on the impact of Ebola on education in Sierra Leone followed by a thorough review of literature on the use of Education Management Information Systems in longitudinal analysis and the use of such longitudinal data in education planning and programming. However, the review found that information on these topics is limited. The details of this review are presented in Section 2. Section 3 goes into the detailed methodology used in carrying out the project. Section 4 tries to answer the research questions by presenting the findings from the analysis of enrolment trends over time and talks through the database structure and dashboard. Section 5 discusses the initial feedback from stakeholders and the ways in which this project has already been put to use. Section 6 provides a brief conclusion to the report.

2. Literature Review

Following the research objectives and the research questions, the literature review was divided into two broad areas: the first being a rapid review of the discussions around the impact of Ebola on education and schooling, and the second focusing on the use of Education Management Information Systems (EMIS) in longitudinal analysis to inform education planning and programming. The review looked at a number of different dimensions, including access to schooling, girls' education, quality of teaching and learning, and education management and governance.

Overall, the literature is hampered by the problems of data consistency discussed elsewhere in this paper, but it generally concludes that the Ebola pandemic affected girls' enrolment negatively, with varying magnitude. In Sierra Leone, using data from a randomised controlled trial, when schools reopened after the outbreak, enrolment was found to have fallen by 32% in high disruption villages (↑Bandiera et al., 2019). Similarly, ↑Smith (2021) compared change in dropout rates in Sierra Leone and Guinea before and after the outbreak, using public survey data. For Sierra Leone, there is a slight increase in girls' dropout rates, though boys show a greater increase in dropout rates. In Guinea, at primary level, girls' dropout rates increased by almost 18%. However, in the secondary level, the dropout rates reduced. In Liberia, even though qualitative interviews with stakeholders indicate that enrolment quickly recovered (↑Santos & Novelli, 2017), actual enrolment data showed a decline from 93% to 85% in primary level (↑The World Bank, 2020).

There are many barriers that prevent girls from re-enrolling back into schools, even after schools reopen. ↑Bandiera et al. (2019) observed that girls shifted to income generation activities to compensate for lost household income. 'Ebola orphans', who lost their parents during Ebola, and pregnant girls face stigma in returning to schools (↑Hallgarten, 2020). In Liberia, parents kept children away from schools for fear over Ebola and lack of understanding of how schools would prevent an outbreak from happening (↑ACAPS, 2016). Additionally, confusion over reopening dates also led to low attendance in Liberia.

The literature review for the second question focused on looking at the data architecture used in different contexts, effective data architecture for a sustainable EMIS, and looking at other developing countries to see if any with similar contexts have successfully created a longitudinal census database. The literature usually takes a deficit view, highlighting the challenges faced by low- and middle-income countries in collecting regular, high-quality data. Another key theme within the literature was how this data should be used to improve learning — for example, within the SABER framework, ↑Abdul-Hamid (2014, p. 19) depicts the ideal use of an EMIS as a dynamic 'information cycle', with feedback loops informing subsequent data collection exercises. This literature highlights three crucial aspects to improve the systems: involving key stakeholders in the process (↑Cassidy, 2006; ↑Powell, 2006); understanding the needs of intended users (including clients) for it to be successfully implemented; and creating a demand for information and information use (↑Hua & Herstein, 2003) at various levels of the service chain, including schools.

Table 1 below provides a summary of the studies included in the literature review. Given the practical nature of this paper, the most useful literature was a recent publication by Van Wyk and Crouch for UIS / GPE (↑van Wyk & Crouch, 2020) that provides insights on how to improve the EMIS at the technical level, including recommendations on EMIS architecture and the way in which data is collected, stored, and reported. It also tries to specify a standardised default EMIS platform and what the countries, in consultation with the development partners, can do to develop such systems.

Table 1. *Summary of literature reviewed.*

Year	Author	Area	Type of paper	Overview	Notes and key points	Relevance for this research
2020	Van Wyk and Crouch	International	How to guide	Explores the key practical steps required to build / manage a functioning EMIS.	The paper provides an excellent guide to how school system databases should be set up and structured, with clearly set out recommendations.	The buyer's guide section of the report provided insights on what buyers should be looking for when thinking of having an EMIS in their country. The following key elements provided the project team with key principles when developing the longitudinal database: an architecture that allows for easy integration of other data sources; a dashboard that allows for new reports of varying depth to be developed; use of a platform that is sustainable from both cost and maintenance perspectives; and finally, technology that is context relevant.
2014	Abdul-Hamid	Low- and middle-income countries	Descriptive study	Discussion on whether EMIS are fit for purpose.	This explains the rationale for an EMIS benchmarking tool, provides an overview of current data-related demands to improve education, and explains how an EMIS meets those data demands. It then articulates how	The paper provided a list of key reasons why EMIS fail including issues like: Misalignment of Activities and Unrealistic EMIS Goals; Institutionalisation of the EMIS; Sustainability Challenges Resulting from Inconsistent Leadership; Missed Integration Opportunities; and lack of use of EMIS at local levels. The project used the findings to ensure that this work for the Government of

					the guiding principles behind an EMIS drive actionable policies.	Sierra Leone (GoSL) takes into account all these key issues and mitigates them at the development stage.
2014	Nayyar-Stone	Pakistan	Descriptive study	Explores the use of EMIS data directly at the source as a means to help bypass problems in the national system in a decentralised setting.	The authors found that improvements of school infrastructure contributed to a decrease in low-performing schools, decrease in teacher absenteeism and a reduction in repetition rate.	This paper provided some examples on how national and sub-national reports generated from the EMIS can help policymakers solve critical issues with service delivery. The project used the concept of providing the right level of information at the right frequency to different decision-makers at varying levels (national, district, chiefdom, school) of the education service delivery chain to make the EMIS relevant for all stakeholders.
2010	Lassibille	Madagascar	RCT	How much do the workflow-enhancing interventions alter the behaviours of service providers and schooling outcomes? How do the interventions differ in impact when aimed at the school, subdistrict, and district levels? Do the interventions	The authors found that the perception of school management improved in treatment schools (37%) after two years. This was considerably higher than control schools where only 15% were considered well managed. However, the impact on student learning was	This paper also provided useful insights on how to make any data system (in this case an EMIS dashboard) relevant by ensuring that the reports that are relevant for a decision-maker are available to them at the right time. The project also used the concept of dashboards having the ability to drill down on data at all levels to better understand the issues as needed.

				have spillover effects?	small and not statistically significant. The authors conclude that changing service providers is both costly and requires a lot of effort, therefore two years may have been too short a time frame.	
2006	Cassidy	Argentina, Brazil, Colombia, Chile, Guatemala, Mexico	Descriptive study	Discussion of EMIS.	The authors concluded that while there are some promising examples of EMIS being used in the Latin American region, specifically in terms of data collection and maintenance, it is not presently able to deliver all that is needed.	Again, this paper provided useful insight for the project on why EMIS systems fail to deliver with specific focus on lack of internal (within multiple waves of EMIS) and external (with other education data sources) integration and how this could be avoided at EMIS development stage.
2006	Powell	Nigeria, Mozambique, Ghana, and Bangladesh	Descriptive study	Discussion of national EMIS implementation.	The authors found that in each of the four countries EMIS was being utilised during the policy process with varying degrees of success.	Though the project did not look into developing any new tools for data collection, this paper provided important information on how EMIS fail to deliver because of poor data collection processes (including poorly designed tools, poor data

					They conclude that EMIS has an important role to play in decentralisation and that effective links between data collection and utilisation are vital.	collection and lack of data credibility). This paper reiterated the importance of data collection processes to ensure success of education management systems. This highlighted the importance of engaging with stakeholders on how they will have to think about these processes in the long term, as we have discussed further in the body of the report.
2004	Ellison	Gambia Cambodia	Descriptive study	Discussion	This paper highlighted examples of the Department for International Development (DFID) work to support use of EMIS in The Gambia and Cambodia. It advocates for the introduction of an annual sector performance review, annual poverty impact analysis using EMIS data, and the use of agreed sector performance targets.	This paper reiterated the importance of data in improving decision-making in education service delivery and helped the team further understand the importance of this project for countries like Sierra Leone.

2005	Murimba	Southern and Eastern Africa	Descriptive Study	High-level overview of large-scale EMIS initiatives undertaken through the Southern and Eastern Africa Consortium for Monitoring Educational Quality (SACMEQ) programme.	It was concluded that SACMEQ enhances the quality of information systems in countries. The author provides examples such as Botswana where the capacity for the production of an accurate and up-to-date EMIS was aided through SACMEQ. This data has since been used as a tool for diagnosing the education system's performance and for planning the development of the system.	The principles of data integration within the country, between different data sources as well as between countries, were used to think through how education data integration can be achieved in Sierra Leone. Furthermore, the paper proved beneficial in understanding the usefulness of having clearly written principles (operations manual) for developing, operating and regularly updating a data system.
2003	Hua and Herstein	Low- and middle-income countries	Descriptive study	Explores the relationship between effective policy provision and system monitoring through data and information.	This paper provided a good entry point for understanding EMIS and provided useful examples of EMIS, implications of their use for educational management, and considerations for data integration.	The project used this paper to understand how to communicate the importance of different technical principles (integration, data validation, data transformation and management) to policymakers to help them understand how applying these principles effectively during EMIS development will make their job of decision-making easier in the long term.

2003	Sultana	Palestine	Descriptive study	This study examined the usefulness of EMIS for the rational organisation and management of an educational system.	This study found that EMIS have been successful in providing a database where questions about the education system can be answered. They also conclude that EMIS provided vital support in planning at the school level, for example, catchment areas and efficient use of resources. In addition, they found that EMIS are most useful in the first five years of educational development planning.	This study was particularly relevant for ongoing work of using data for catchment area planning and schools approval in Sierra Leone. The study provided the research with important information on how to use school-level data to streamline decision-making, especially decisions related to where new schools need to be developed / opened.

In addition, the project drew on policy reports from Sierra Leone, including the Annual School Census reports, and reports on the data context such as the 'Open Data Readiness Assessment Prepared for the Government of Sierra Leone' by the World Bank. These documents were used to understand key challenges faced by the Government in carrying out the ASC activity. The reports also provided insights on how the government uses the data for education planning and programming. The Open Data Readiness Assessment proved beneficial in understanding internet and smartphone penetration in Sierra Leone and how any changes to the ASC data architecture should keep the internet bandwidth constraints at its core.

Finally, in undertaking the literature review we also sought literature highlighting the experience of similar projects to reconstruct historical systems in other countries. However, such examples could not be found, with literature more frequently concerned with combining datasets across different sources and sectors, than longitudinal compilation of a single source. This gap therefore further increases the importance of carrying out this project and documenting the processes and the lessons learned to help other countries going forwards.

3. Methodology

3.1 Research framework

This paper has two aims. These are based on the construction of a combined database containing school-level Annual School Census (ASC) data from 2015 to 2019. Because the ASC datasets were not provided in a uniform format, the first step was to transform them into a consistent format. This also involved cleaning and reconstructing the data from the ASC into a single, easy-to-query database.[3] The team then moved on to answering the research questions.

The first aim of the paper is to examine the main trends in education enrolment dating back to the pre-Ebola period (RQA1 through RQA4; see Section 1.3). To this end, the paper largely analyses secondary data using the integrated database. It also includes analysis of Multiple Indicator Cluster Surveys (MICS) data for specific research questions.

The second aim of the paper is to provide recommendations on how to use and manage an education system's data architecture (RQB1 through RQB3). To do so, the study drew on existing literature, stakeholder conversations, and capacity review of Sierra Leone's existing data system together with the learnings from the process of creating a multi-year ASC database.

An overview of the data sources and methods for each research question (in Section 1.3) is summarised below.

Table 2. *Research framework.*

Research question	Data sources	Methods
RQA1. What lessons can be drawn from school-level changes in enrolment following Ebola, which can be applied to school reopening following Covid-19?	Sierra Leone Annual School Census datasets and reports — to create an 'integrated database'.	The approach will look to recreate the database at the district level, using data extracted from the school census reports (for pre-Ebola data) and aggregated from the school-level data (for post-Ebola data).
RQA2. How did enrolment change across the country? Are there any hotspots of incidence?	Integrated database.	The analysis will look at trends across the country using simple time-series plots of enrolment over time. The analysis will also look at the

[3] Since completing this research study, the team has included the 2020 ASC data in the integrated database.

		district-level data to identify spatial variation.
RQA3. Are there any differences by gender in terms of re-enrolment following Ebola?	Integrated database.	The analysis will look at trends over time in the data and how enrolment levels vary across districts.
RQA4. Is it possible to use these patterns, and household survey data, to predict likely areas of vulnerability going forward?	Integrated database; alongside MICS data.	Analysis will include running a Random Forest algorithm on the MICS data to try and identify the factors that are predictive of girls' dropout. The analysis will then look at how these factors can be measured at the school level, to identify at-risk schools.
RQB1. What does the optimal education data architecture look like, and what are the key initial steps to take towards building such an architecture, given past experiences and the constraints on the field?	Rapid literature review; informant interviews.	The analysis will draw on the wider standards for database management, including enforcing relationships to ensure database integrity; the use of identifiers; and discussions with database users and designers.
RQB2. How can we demonstrate how the Annual School Census (ASC), built with a well-planned new data architecture, can help the government in improving education planning by using existing ASC data?	Integrated database.	The analysis will use the merged database to create an interactive dashboard of curated messages, to allow policymakers in Sierra Leone to delve deeper into the data. This will be handed over to the Government of Sierra Leone (GoSL) and integrated into the education datahub.
RQB3. What are the key components / data frames for this system — which currently exist, what are planned, and how can they be linked?	Key informant interviews.	Key findings, processes, and learnings from the project will be discussed with key informants and data users in Sierra Leone, to identify possible components of a unified data architecture.

3.2 Data

The project involved the construction of a consolidated database incorporating school-level ASC data from 2015 to 2019 as the underlying data source for all the research questions. The team partnered with the Ministry of Basic and Senior Secondary Education (MBSSE) to convert historically fragmented ASC datasets into a single fully integrated database with an interactive dashboard.

The research team took the following approach. Firstly, this involved understanding the contents of the datasets. This is described in more detail below. Next, datasets were cleaned, combined, and used to populate the integrated database. Then, once the database was developed, an interactive user interface (UI) was developed to carry out required analysis and show the results in a format that conveys the key messages to the policymakers and helps them with education planning and programming. The latter two processes are explained in more detail in the results section as an output to guide other countries who may also wish to undertake this process. Finally, the database was used to find the enrolment trends and any other long-term trends in education in Sierra Leone. These results are described in detail in Section 4 of the report.

Table 3 describes the key data sets used to construct the consolidated 2015–2019 ASC database.

Table 3. *Summary of data sources, data format, and data outputs.*

Available data		
Data	**Current format**	**Planned outputs**
ASC 2015	MS Access	Integrated database
2016	MS Access	
2017	CSV (tables)	
2018	CSV (wide)	
2019	CSV (wide)	

Understanding the contents of the datasets

In first understanding the datasets, one key discovery was that a dataset that was labelled as 2014 and expected to be data from before Ebola, was actually from 2015 and was from the first term after Ebola. This is discussed in more detail in the challenges and limitations section. This meant the project had to look for alternative ways of analysing the before- / after-Ebola education trends.

For this reason, the data for this project was divided into two categories:

a) Post-Ebola detailed data which was compiled into a new database using the 2015–2019 ASC raw datasets;
b) Pre-Ebola summary data which was taken from the 2010 / 11, 2011 / 12 and 2012 / 13 ASC reports.

Since the pre-Ebola data was only available in word format, the data had to be extracted from relevant tables within the annexes of the report, which limited the analysis to district level only rather than school level as was initially planned.

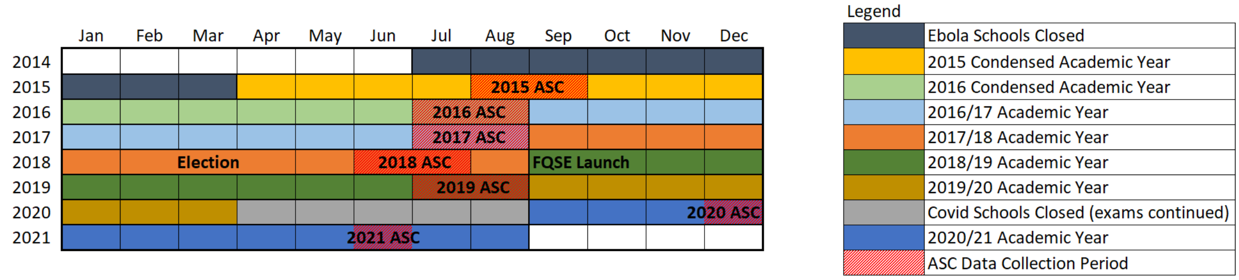
The ASC datasets were not provided in consistent formats. As explained in Table 3 above, the 2015 and 2016 datasets are in Microsoft Access (MS Access) tables (i.e., with separate tables for classrooms, disability, furniture, facilities, students, teachers, textbooks, and so on). The 2017 data are in a similar table structure, but each saved separately in Comma Separated Value (CSV) files. The 2018 and 2019 datasets are in CSV files each structured in one single dataset at the school and teacher level with more than 100 columns each. This mismatch of formats, lack of unique identifiers on which to match the schools across years, duplication in school records within and across datasets, and lack of consistency in which schools, districts, and chiefdoms were named in the datasets from different years was a significant challenge to overcome in combining into one database.

Given the structure of the data, where some of the data is one-to-one (one record per school, for example, enrolment per class) and some of the data is one-to-many (multiple records per school, for example, details of teachers per school), the separate tables format of the 2015, 2016, and 2017 data is significantly more efficient than the wide format of 2018 and 2019[4]. Therefore, this is the structure that is used in the combined database and requires consideration of the level of data aggregation for each element (such as child / grade / school level) which is discussed further below.

For the integrated database the school-level data from the 2015, 2016, 2017, 2018, and 2019 ASC were incorporated into the database. Figure 1 below shows the data collection periods and the academic year in which these occurred. The panel on the right-hand side also shows the key events affecting the education sectors during that period.

[4] A single table is good when data is one-to-one. When you have many thousands of rows and columns of data and the data is one-to-many, multiple tables are more efficient for storage and analysis, as this reduces duplicate data.

Figure 1. *Timeline of ASC within school and calendar years. Given that the project had access to 2015–2019 raw data, it was possible to understand the data in detail compared to 2012 and 2013 pre-processed data. The details of the 2015–2019 data are provided below.*

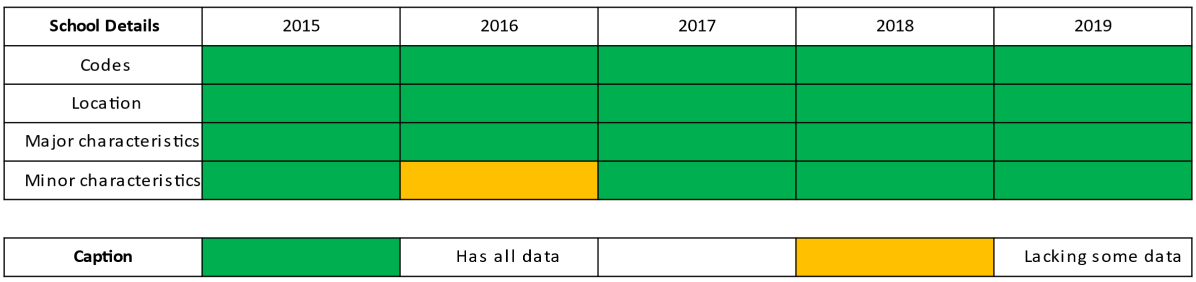


At a higher level, the ASC data for 2015 to 2019 can be categorised into three main strands: enrolment, teachers, and school facilities. However, the exact data that is collected each year varies, not just in terms of the information but also the level of disaggregation at which this is collected. The extent of consistency and coverage of the data in each ASC is shown in Figures 2 to 5, below.

School data

Figure 2 below shows the level of data completeness on school related indicators across years. The green cells show that for all five rounds of ASC the schools had all the information except for some missing data on minor school characteristics for 2016 data. Here major characteristics are defined as, for example, school level, owner, approval status, or bank account, and minor characteristics are defined as other information such as School Management Committee data. These categories were developed in consultation with the MBSSE and EdTech Hub team.

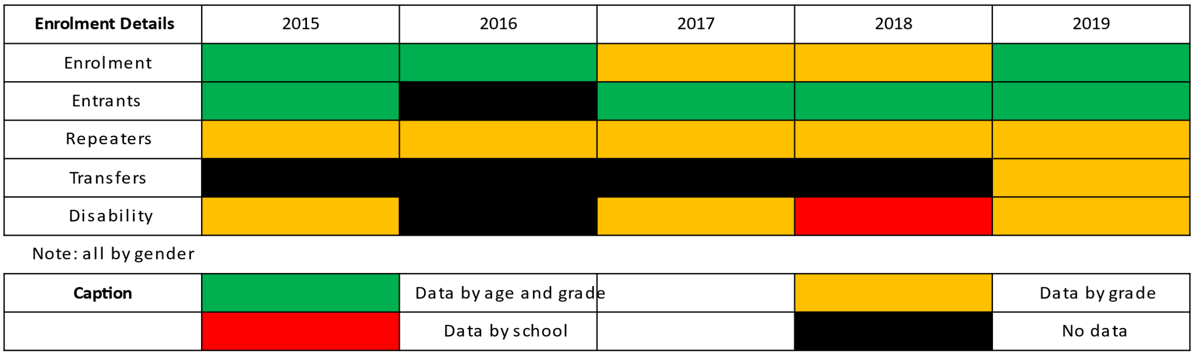
Figure 2. *School-level data details by ASC. Note: the row labelled 'Caption' provides a key detailing cell colour coding.*



Enrolment data

The enrolment data was found to have more discrepancies across the years compared to school-level data. The black cells show that transfer data was only added in 2019 and was missing in all rounds before that. Similarly, 2016 data did not gather information on new entrants or student disability status. 2017 and 2018 only had enrolment data by grade whereas the rest of the ASC rounds had data on enrolment by both grade and age. Fortunately, all enrolment data was gender disaggregated for all the ASC rounds.

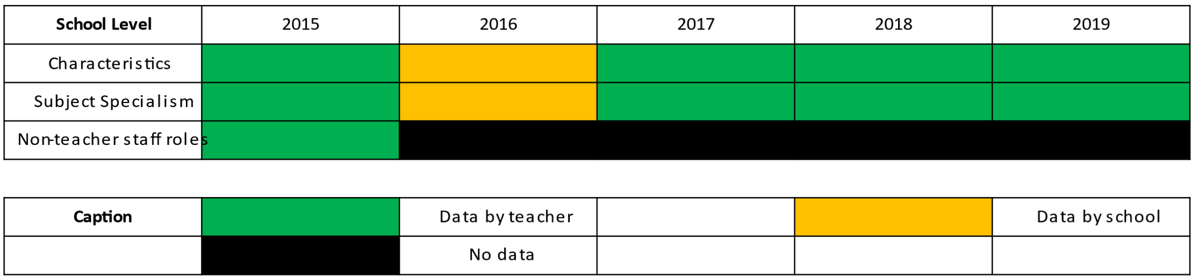
Figure 3. *Enrolment data details by ASC round. Note: the row labelled 'Caption' provides a key detailing cell colour coding.*



Teacher data

One of the key findings when analysing the teacher data was the missing non-teaching staff data for all ASC rounds after 2015. Additionally, the 2016 teacher characteristic data (such as gender, age, years of service) was provided as aggregated school-level data in 2016, whereas for the rest of the year this data was provided for each individual teacher. The same was the case for teacher specialisation data for 2016 compared to the rest of the ASC rounds. This is detailed in Figure 4 below.

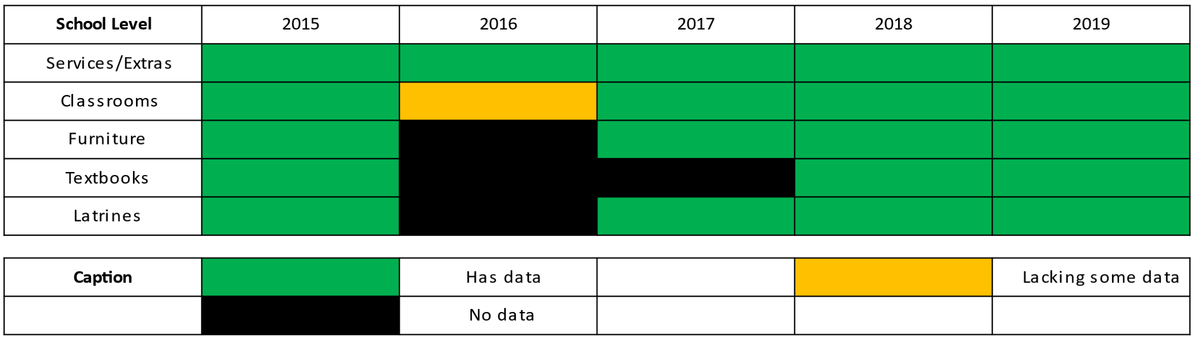
Figure 4. *Teacher-level data details by ASC round. Note: the row labelled 'Caption' provides a key detailing cell colour coding.*



Facilities data

All ASC rounds except for 2016 were found to have relatively consistent levels of data on school facilities. 2016 ASC did not gather information on textbooks, furniture, or latrines in school and was found to have some data missing for classrooms. Similarly, 2017 data did not have any information on textbooks. This is shown in Figure 5 below.

Figure 5. *Facilities data details by ASC round. Note: the row labelled 'Caption' provides a key detailing cell colour coding.*



These five datasets form the basis of the merged database (the project has also added the 2020 data to the merged database since finalising the research study).

Additional data

To this integrated database, district-level data on enrolment was added from the 2010 / 11, 2011 / 12 and 2012 / 13 ASC school years reports. In addition, the database pulls in further data sources for tailored uses: Ebola case data by week and by district from the World Health Organisation; 2017 Round 6 MICS data; high-resolution spatial population data; and high-resolution relative wealth indices. These are detailed in Table 4 below.

Table 4. *Summary of data sources, data format and data outputs.*

Additional data for analysis

Data	Current format	Planned outputs
ASC Reports 2010 / 11 to 2012 / 13	MS Word	Data incorporated at district level into the integrated database
Ebola Case Data (↑World Health Organization, 2016)	CSV	Data incorporated at district level into the integrated database
MICS (↑The World Bank, 2017)	CSV	Python Machine Learning model
WorldPop High Density Population (↑WorldPop, 2018)	Shapefiles	ArcGIS
Relative Wealth Index (↑Facebook, 2021)	CSV with geo-data	ArcGIS, and incorporated at school level into the integrated database

3.3 Research instruments / tools

This study conducts analysis of secondary data. It therefore relied on the products of the pre-existing instruments. This includes the questionnaires for the ASC, and the MICS survey round 6, as well as the methodologies of the WorldPop high density population mapping and Facebook Data for Good's (D4G) Relative Wealth Index.

3.4 Research design

This study adopts multiple methods, largely using analysis of secondary quantitative data. These vary in accordance with the research objectives and are summarised below.

A. First aim of the research

RQA1 to RQA4 use analysis of trends on the integrated multi-year ASC database to compare enrolment changes and Ebola cases over the pre- / post-Ebola period, as well as longer-term education trends. This was complemented by the ability of the database to track schools across time, which meant that even though GPS data on schools was not collected until 2018 the location data could be backdated to allow for mapping and geospatial analysis prior to 2018 as well. From this database, data visualisation tools enabled comparison of these factors across a range of aspects, including inspection of any gendered differences.

In addition, to help answer RQA4, the project explores the possibility of using Machine Learning (ML) approaches to predict dropouts. Such ML approaches were used by the research team in the Girls' Education Challenge evaluation work in other countries. For this research a similar process was applied to see if the ML could accurately predict children at risk of dropping out in Sierra Leone. If successful, this would help target support to schools as they recover from the Covid-19 crisis.

To do so, a Random Forest ML algorithm was applied on the MICS data to identify the characteristics at the child level which were associated with the dropouts among the children in the MICS sample. This was done using the MICS 2017 data; full details of the survey are found in ↑Statistics Sierra Leone (2017a). Other ML techniques were considered including Logistic Regression; Linear Discriminant; K-Neighbours; Decision Tree; and Gaussian-Naive Bayes. However, given time limitations and the nature of the analysis, the team decided to use Random Forest.

Random Forests are based on the divide-and-conquer approach of decision trees generated on a randomly split dataset. The collection of these decision trees is known as the 'forest'. Each tree depends on an independent random sample. In a classification problem, each tree votes and the most popular class is chosen as the final result. It is simpler and more powerful than other non-linear classification algorithms and gets a more accurate and stable prediction than a single decision tree. The disadvantage is that they are more difficult to interpret than a decision tree, where you can easily make a decision by following the path in the tree.

It is worth nothing that this work is exploratory and warrants much more attention than can be included here. Ideally more work is needed to refine the ML model, such as more detailed analysis of the overage data and the other available variables, including ranking districts by the indicators that are highlighted as being important in the feature selection from the ML modelling. It can also be tested using 2020 and 2021 data to further validate predictions.

The analysis used the child's enrolment status, alongside the wider household characteristics and location variables from MICS data, to try and identify factors which are most associated with dropout. Dropout was defined as a child having been enrolled last year, but not this year. These factors were then mined in the ASC database to identify schools with particular prevalence of these factors followed by some basic validation using the enrolment data over time (schools with such trends should ideally show a consistent drop in enrolment if the right factors are identified by the ML algorithm). However, since the majority of the factors in the MICS are household variables, these were not available in the ASC datasets thereby limiting the use of these factors for school-level targeting. The key factors found from MICS that could be used at school level were age-for-grade and the higher grade one dropouts.

The research team used the combined database (2015 to 2019) to identify schools with the highest concentration of older children in any grade. The enrolment trends for all such schools between 2015 and 2016 were observed to validate the predictions, looking at the average enrolment changes for schools with high prevalence of these factors compared to those without. 2015 to 2016 data was used for validation since using 2019 data would have skewed the validation because of the increased enrolment in schools after introduction of the Free Quality School Education (FQSE) policy (but interestingly, there was still far lower growth in enrolment for these schools than the average for a given year).

The analysis focused on the schools with high shares of children that were overage for their grade, and those with a very high share of their enrolment in grade one. There is subjectivity in what to consider as a cut-off for schools with high grade one enrolment versus schools with normal grade one enrolment. If children were evenly distributed across the grades, then it would be expected that 16.6% of children would be in grade one (though population growth may increase this to 17 / 18%). Therefore, for the sake of simplicity a school with 40% of the children in grade one was defined as high risk.

Similarly, defining a cut-off for the overage share was more complex, as the term times run for the calendar year. Here, a child could not be counted as overage if they were two years older than the official age for that grade — while grade one enrolment is aged six, a child can turn seven in that year, so only children over eight were counted as overage.

That given, information about overage enrolment has been included in this paper to highlight the potential for an integrated database to be linked to other administrative databases to help identify and target schools with particular characteristics or with particular needs.

Similarly, the project also uses high-density information on relative wealth, produced by D4G. The D4G data was calculated for a three-mile catchment area around each school and is used to consider the relative wealth of catchment areas for each school.

B. Second aim of the research

For RQB1 to RQB3, the research team drew on their experience in education data analysis and management, as well as learnings from the process of creating the integrated database, frequent and ongoing discussions with stakeholders within the country, and existing literature. To support engagement, this has been accompanied by a data-analysis platform using the integrated database which allows users to investigate the data themselves. This is currently hosted on Fab Data's SharePoint but with ongoing discussions for transferral and ownership by the GoSL.

3.5 Stakeholders

Given the importance of EMIS data in education planning and programming, a wide range of education stakeholders in Sierra Leone were interested in understanding the approach the project has taken to make the data longitudinal and more reliable. Particularly relevant government stakeholders included the Delivery Unit and Directorate of Planning and Policy of the MBSSE; other key government institutions included the Teaching Service Commission and the Directorate of Science, Technology and Innovation. These stakeholders were consulted at each stage of the project and have been involved since the conception, with strong in-country support from EdTech

Hub, who helped facilitate regular check-ins with the Delivery Team in MBSSE. In addition, the project also consulted education development partners in Sierra Leone such as the FCDO, World Bank, and IIEP-UNESCO.

3.6 Ethical considerations

This project did not directly engage with human participants and drew exclusively on the existing data and data collection instruments in Sierra Leone. All data was processed in compliance with Fab Inc. data protection policy and all our staff and partners adhered to our safeguarding policies.

This project followed the ethics requirements of the MBSSE. Prior to conducting the research, the team also asked themselves the following questions to ensure that the project constituted ethical research and to determine if it was ethical to conduct this research during an emergency.

Is this research topic of practical use?

This topic is clearly of interest to policymakers, who have been involved in the project throughout. This project can also help to improve equity with its findings.

Is there a risk to participants?

There is limited risk to participants as no fieldwork is required. No interaction with vulnerable children or adults is required.

Will implementing these procedures or methods impact on confidentiality or privacy?

No, this will work solely with the current administrative data that has been shared with us by the Ministry previously and is stored securely in line with data protection policies.

Do safeguarding procedures need to be adapted or updated for this research?

As there is no direct contact with participants, there is very little safeguarding risk.

The team have worked closely with colleagues in Sierra Leone throughout the process in order to review findings and to consider the language and dissemination of findings.

3.7 Challenges and limitations

The project faced a number of challenges and limitations, mostly centred around the consistency of data and challenges within datasets. One particular challenge concerned the before- and after-Ebola analysis.

During part of the data understanding and cleaning exercises, the team noticed that the dataset that was labelled 2014 had time-stamps for the data collection from 2015, and was in fact data from the first term post-Ebola (this is therefore referred to as the 2015 data throughout this report). As such, the research questions relating to the school-level database were updated to look at changes between 2015 and 2019, and the pre- / post-Ebola analysis was conditioned upon finding the data before Ebola. The team managed to source pre-Ebola ASC reports, which contained annex tables of enrolment at the local council level by gender. With some restructuring, this then allowed for some investigation of the before- / after-Ebola enrolment.

However, this meant that school-level changes over time could not be analysed, and the team had to rely on district-level averages. It also meant that the team were limited in the geospatial analysis, as many of the possible spatial extrapolation methods relied on more disaggregated data. While a limitation, this also highlights the benefits of our second strand of research, which looked at creating a centrally managed database — rather than various files containing data that are distributed on local computers with no version control and no change tracking.

Another significant challenge was that creating a usable data architecture took far longer than expected. Despite the team's vast experience with the data Extract, Load, Transform (ELT) process, the extent of the differences in the data collection, processing, and storage between 2017 and 2018, and particularly the extent of issues that arose within the 2018 data, took much longer to resolve in order to reach a high-level of school matching between the years. This showed that data quality is not always improving, and that the amount of time required to resolve some of these challenges can be significant. Again, while this was a challenge, it also informed the work in setting out a step-by-step guide for others to follow for implementing similar databases in other contexts.

An additional limitation was the inconsistency of data collection. Of greatest importance is the schools that are included in each year — the numbers changed, but often it was unclear if this was due to changes in coverage or new schools being founded (a genuine increase). Review of ASC reports shows that census activities do not necessarily cover 100% of the schools in the country (as is desired from a school census). For example, the 2018 ASC report cautions that the increase in the number of schools in the data '*should not be considered as an actual increase in the number of schools but rather improved coverage. It could also mean that in the previous school censuses some school heads were not returning their questionnaires back to the office for processing*' (↑MBSSE, 2018, p.8).

The inconsistent collection of different aspects of the data each year was a limitation that also reduced the ability to analyse and compare all data across all years. In some of the cases where data fields were collected at different levels (for example, at the grade or school level) in different years, the intention was to retain as much information as possible by keeping the most disaggregated level as well as populating a more complete aggregated level, but this required us to do some coding to extrapolate data and disaggregate it. This is not ideal in terms of the data population as these cases simply didn't have all the same information that was available in other years, but it maximises the potential data use in these cases of inconsistent data aggregation.

Another limitation was the sheer scale of the data that is collected. As an example, the 2019 school-level dataset has more than 1200 variables. This scale and the variation of data means that decisions had to be made in terms of which variables are considered 'core' information for inclusion in the database. The breadth of options for presenting this also meant that curation was required for the database to become usable.

The final limitation was the inconsistency of reporting in the database across years. The lack of change tracking meant that where information changed between years it was not clear if this was updated data or mistakes in data entry. While this does not stop the data from being used, it did mean that analysis at a very local level could often display strange results, which ideally would be flagged and queried as part of the data assurance process as this data is collected.

4. Results

As previously mentioned, the underlying data for all study objectives and research questions was based on the construction of a consolidated database incorporating school-level Annual School Census (ASC) data from 2015 to 2019 as well as an associated dashboard. Creating a usable database took longer than planned, especially when it came to reaching a high number of schools matching across the years. As a result, the team developed a ten-step guide to facilitate data utilisation.

This section presents the guide in terms of how the process was carried out to meet the specific needs of Sierra Leone's education data, but is also intended to summarise how this could be replicated in other countries that do not have a school census database with consolidated school-unique identifiers across years.

The database created was then used to examine the enrolment trends and any other long-term trends in quality education in Sierra Leone in terms of inputs such as teachers, facilities, and infrastructure. Before diving into the research questions, we present these findings.

Finally, the section presents the findings as responses to primary and secondary research questions. These are organised by research objectives A and B (see Section 3.4).

4.1 Ten-step guide to enable single, easy-to-query access to data

The process of integrating Sierra Leone's ASC data from 2015 to 2019 and creating the user interface is explained below.

Ten steps to enable easy use of data
Step 1: Reviewing the data
a) **Reviewed the 2015 ASC**. The data was in MS access format. All the tables in the database were reviewed to understand the relationships between the data. The data and the relationships were documented for comparison with data from other years. b) **Reviewed the 2016 ASC**. The data was also in MS access format. Again, all the tables were reviewed to understand the relationship between different tables (such as school characteristics, students, teachers). Here an additional step was added to where data and tables were also compared with 2015 data to understand how the data could be matched and how much data was missing in either of the datasets. 2016 data was found to have significant missing information.

c) **Reviewed the 2017 ASC**. This data was provided in the CSV format with each file being a separate table for school characteristics, such as student enrolment, teacher information. Though 2017 had most of the information that allowed it to be matched with 2015 data, 2017 data was missing some information on school locations. However, having school IDs allowed for generating the location code and matching them to the 2016 and 2016 location codes.

d) **Reviewed the 2018 data**. This data was a large shift away from the previous three years. The data was in two wide CSV files (School and Teachers). The data was arranged in a 500+ column-wide single table. Between the 2017 and 2018 rounds of ASC, the administrative boundaries for districts and chiefdoms had also changed, introducing new chiefdoms and districts that were not present in previous ASC rounds. Since school IDs were based on districts and chiefdoms, this meant the 2018 IDs were not viable for matching to previous years any more. The change in administrative boundary also resulted in many schools incorrectly reporting the district and chiefdom, which made matching more difficult. Thus, the project used location estimation using GPS coordinates to match the school from 2018 to schools with 2015 to 2017. This helped in matching more than 90% of schools. However, this wave of ASC was found to have significant duplicate entries (primarily due to structure of the datasets). A number of schools in the dataset were missing school IDs and had codes like 99999999 — which was not an issue in the previous years. However, with location estimation we were able to match the school with the previous dataset and use the school ID from the previous year.

e) **Reviewed the 2019 data**. The data was in a similar format to 2018. However, the dataset was even wider with many more columns, up from 500+ to over 1900. The data did not contain the same level of duplication that the 2018 data did. However, it still had the problem of school IDs changing and duplicated school IDs. Again, location estimation using GPS coordinates was used to match schools to previous years.

Step 2: Map out the initial high-level data structure

a) Having **reviewed each of the ASC datasets**, the project had a good idea of the relationships between the data and therefore the project mapped out a high-level overview of the tables that were needed including a map of how these tables relate to each other.

Step 3: Produce and populate the core level 1 tables

a) **Level 1 tables** are the tables that have no dependencies. In other words, other tables depend on them, but they do not depend on any others. Examples include: Districts, chiefdoms, school type, grades.

Step 4: Import the first raw data

a) Since **2015, 2016 and 2017 data** were already in a suitable structure, import routines were run first to **import this data.** However, the data being in different formats — Access and CSV — required us to use various routines to include conversion of all Access tables to CSVs before they could be imported in the new database.

b) Once the data was imported, **validation checks were carried out on the data.** These included checks of duplication and checks for missing values in the key columns (such as school name and district name).

Step 5: Create unique ID mapping table

a) In order to track schools over time a unique ID needs to be assigned to school that doesn't change over time. Therefore, **the existing school IDs could not be used**.
b) To solve this **a new unique code** format was created and all the tables were populated using the new unique ID. New code format was mapped to the school IDs year by year from 2015 to 2017, and new code was only generated where the school did not exist in the previous year.

Step 6: Determine the core data for use in the final structure

a) Based on the high-level data structure that was mapped, and the review of the data in each of the years, **data to be imported** to each of the tables was defined.
b) **Which tables needed to include change tracking** was also defined so that differences over time could be reported. These were set to be: districts, chiefdoms, councils, and schools.

Step 7: Iterate through the designed structure, create the tables, and populate

a) Finally, the **tables were created** and data relationships were added.
b) Code was written to transform the imported raw data into the required form and **populate the tables**.
c) **Validation checks were carried out on the imported data**[5] against the raw data to ensure that import did not alter the results.

[5] There were multiple strands to the validation. Firstly, once the raw data was imported, queries were tested on all data (such as enrolment, teachers, facilities) to calculate the various counts and compare to raw column data. This allowed the team to ensure that there were no issues with the imports and that data had not been changed due to differences in datatypes or due to the import processes themselves.
Queries were then written to pivot the data into the required form. Since these pivots considerably changed the structure of data, the pivoted data was verified by comparing the counts and sums of the raw data with pivoted data to ensure these also tallied with the previous steps.
Outside of the calculation steps there were multiple stages of secondary validation. For example, at the teacher level issues were found with the teacher ages — teachers being too old, for example. In these cases, data was reviewed —in the previous example the age was set to be blank (unknown).
At the school level, issues were observed with GPS coordinates at latitudes and longitudes of 0,0 which is clearly a data entry error, so these coordinates were removed as they are invalid.

Step 8: Clean and import the additional raw data into the structure
a) 2018 data was de-duplicated and cleaned, removing schools that were obvious duplications. b) Merged the 2018 data with the 2017 data and assigned the new unique code. This was done in multiple stages, starting with the strictest matching criteria and reducing down. At each stage of merging, it was ensured that there were no duplicates. New unique codes were assigned records that couldn't be merged, i.e., new schools. c) Imported the 2018 data into GIS software. To correct administrative area errors, spatial analysis was carried out using the new latitude and longitude data to assign the correct chiefdom, district, and region. d) Imported the clean 2018 data into the database. e) Replicated the process for 2019 data, but this time merging to the newly cleaned 2018 data.
Step 9: Iterate through the designed structure and populate
a) **Followed a similar process to Step 7** (without the table creation) to iterate through the structure, and populated the data.
Step 10: Visualisation
a) Imported the data from the new structure into **business intelligence software** to better visualise. b) Created **calculations** for measures that relate to multiple variables, such as Pupil-Teacher Ratios. c) **Visualised the data in dashboards**[6], and continued to improve utility based on discussions with stakeholders.

[6] Creating the visualisations was a fairly involved process. Firstly, the SQL database was connected to Power BI and the tables and data were imported. Power BI understands the relationships that were created in SQL server and is able to re-create this structure within the Power BI tool.
Once this is done the dataset was then configured for report building. There are many columns and pieces of data within the database that are present for database operation but which are not relevant to people who are building reports. These columns were hidden from report-builder view. The data columns that are useful for the report builders were then configured for how the reports were to be displayed, such as setting the formatting, Power BI specific datatypes (for example, locations) as well as how the data in the columns should be summarised by default when being used in reports.
At this point visualisations were generated for a good amount of the base data, but it was still necessary to specify how to combine these data points for complex calculations. For example, to calculate the pupil–teacher ratio, calculations were to be carried out in Power BI to sum the total number of teachers vs the total number of students, and these calculations needed to work across all filters the user may select. Describing how to write Power BI calculations is beyond the scope of this report.

4.2 Overarching key trends in Sierra Leone's ASC database (2015–2019)

Before diving into the research questions, below are some of the key findings that did not directly relate to the research question, but were worth highlighting as they emerged as key trends from the 2015–2019 merged database:

- Enrolment:
    - The huge extent of the increase in enrolment in 2019 as a result of the Free Quality School Education (FQSE) policy is especially apparent, particularly in comparison with the slow progress in the years beforehand.
    - This increase is evident across all districts and school levels, in particularly for approved schools, and emphasises the additional strain this has placed on the education system.
    - Repetition rates have varied year to year but show a downward trend to a five-year low in 2019.
    - Significant strides have been made in terms of gender parity, particularly at secondary level, but significant spatial distinctions are evident, with noticeable disadvantages for girls in the north of the country and the opposite in the south.
    - The number of children with disabilities has increased significantly, particularly between 2017–2019. This has included increases in the total number of children, as well as their share of total enrolment, across all disability types.
- Teachers:
    - There was a noticeable increase in the number of teachers in 2018, which actually precedes the increase in enrolment in 2019 as a result of launching the FQSE policy. However, the 2018 ASC data collection was conducted in July — after the election of President Bio on an education platform in March — which might explain this timing, with the new teachers pre-empting the increase in financing to the sector.
    - The increase in teachers in 2018 was particularly from unqualified volunteers — likely anticipating future additions to the

government payroll — which has reduced the overall share of qualified teachers between 2013 and 2019.

- Pupil–teacher ratios are much lower than pupil–classroom ratios, limiting the ability to efficiently utilise the education workforce. This also includes secondary level, where there are significant and worsening shortages of core subject specialists.

- Facilities
  - The numbers of schools and classrooms has not increased at the same pace as the number of students. As a result, the need for additional classrooms to meet targeted ratios of pupils-per-classroom has increased significantly in 2019.
  - There also remain significant infrastructure challenges in terms of classroom repair needs, latrines and water access, particularly at the primary level.

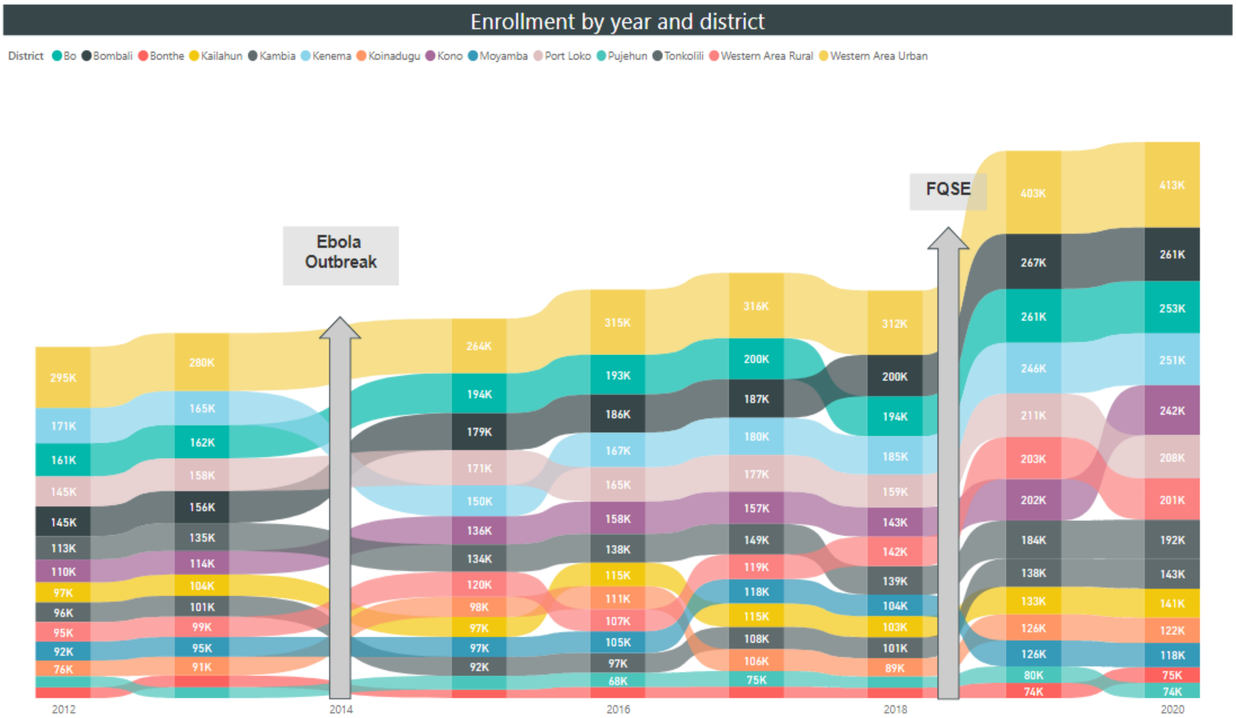
4.3 First aim of the paper (A): RQA1 to RQA4

The 2014–2016 Ebola epidemic in West Africa was the largest ever recorded, affecting Guinea, Liberia and Sierra Leone. Schools throughout the country shut down in July 2014 and remained closed until April 2015. Given this, and concerns over increased dropouts, especially for girls, the research looked to see how enrolment has changed over the last nine years, with a focus on both the short-term impacts (enrolment changes in 2012 / 2013 school year and the 2015 school year); and long-run recovery (looking at 2011 / 2012 to 2019 / 2020).

The analysis revealed that while the school closures from Ebola did lead to a fall in enrolment in some districts, these falls were relatively short-lived, with enrolment numbers broadly recovering by 2016 and increasing noticeably over the years afterwards, with a huge jump in 2018 / 2019 across all districts following the introduction of the Free Quality School Education policy. This is shown in Figure 6 below[7].

[7] Note that Figure 6 sorts by the size of the district each year. This is important to note in the case of Moyamba for example, where its relative ranking within districts falls between 2018 and 2019, but a closer look at the values shows that enrolment still increased from 104,000 pupils in 2018 to 126,000 pupils in 2019.

Figure 6. *Changes in district enrolment 2011–2020.*

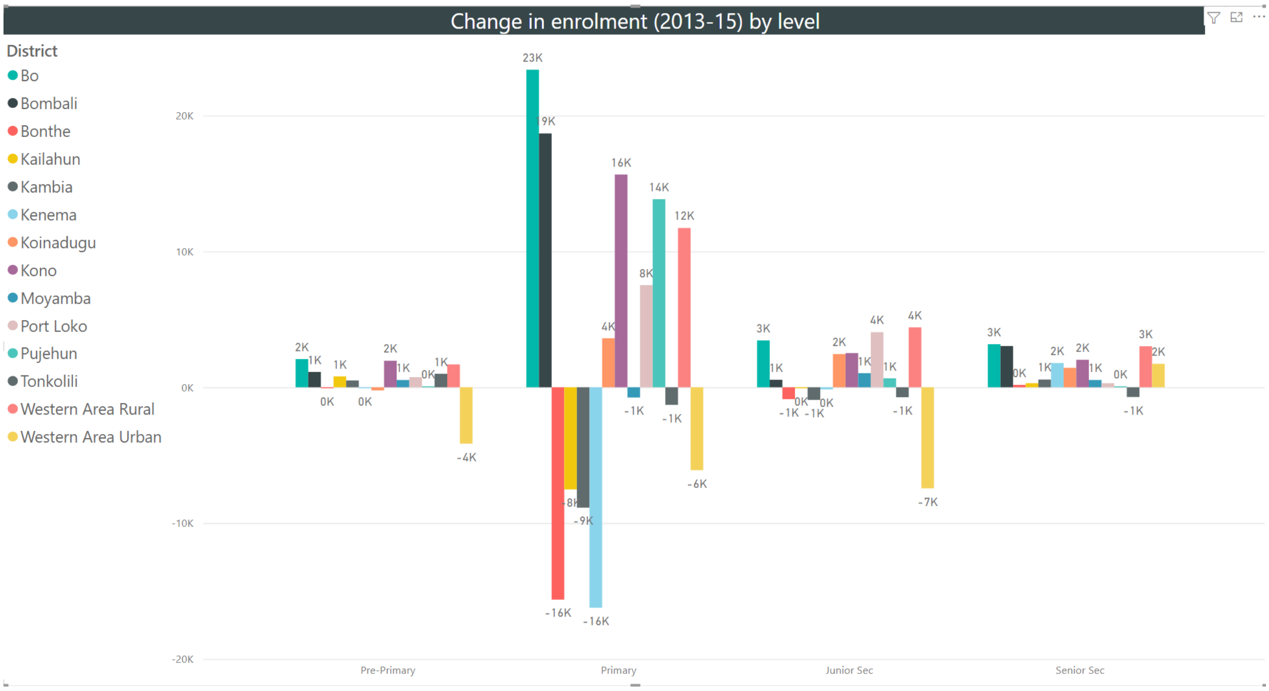


The individual sub-questions will be investigated in detail first, before returning to the headline question (RQA1) and the lessons that can be learned at the end of this sub-section.

RQA2. How did enrolment change across the country? Are there any hotspots of incidence?

As shown in Figure 6, while the country as a whole showed an increase in enrolment numbers in 2015 compared to 2012 / 2013 (an increase of 70k children, or about 4%), this masks some local variation. Figure 7 shows a detailed breakdown of enrolment by district and level of schooling to understand local variations. When looking at the absolute number, more children were found dropping out of primary schools compared to secondary schools. However, when looking at the percentage changes, it was found that the changes are equal across levels.

Figure 7. *Changes between 2012 / 2013 and 2015 by school level.*



The greatest declines are found in Western Area Urban, Bonthe and Kenema, where enrolment fell by approximately 15,000 students across all the levels. However, the patterns within these vary, with Western Area Urban seeing falls across both primary and junior secondary, while in other districts the declines were concentrated in primary schools. Only one district, Tonkolili, showed a (minor) fall in enrolment at senior secondary schools.

One note of caution is that the 2015 dataset includes a noticeably higher number of schools than the 2012 / 2013 dataset, with coverage increasing from 7,967 schools to 9,227 schools — an additional 1,260 schools. As discussed in Section 3.7, this is a complex challenge[8].

The research also looked at the Ebola case data to estimate the severity of the outbreak for each district, which could help further understand the before- / after-Ebola enrolment trends at district level. The number of recorded Ebola cases during the pandemic varied noticeably across districts — some districts, such as Bonthe and Pujehun, recorded less than 60 cases in total while others, including Port Loko and the Western Area Rural and Urban, recorded more than a thousand.

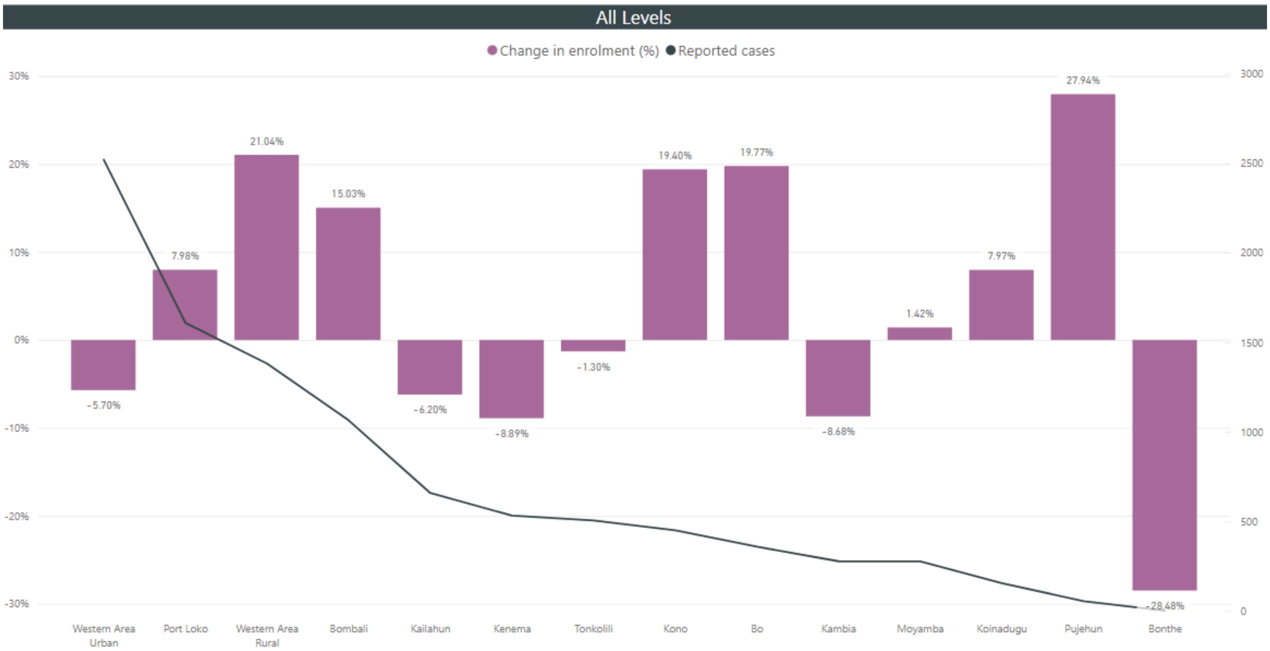
The emergence, along with the severity, of the virus was not uniform either as a handful of districts experienced their most severe outbreaks in late July of

[8] Within the 2015–2019 combined database, workarounds can be identified such as restricting analysis to only schools that are matched and identified as existing in both years of interest — but this is not possible with the pre-Ebola data that is not available at the school level. Moreover, it is likely that the missing schools are smaller or more remote, which would reduce the accuracy of simple extrapolation.

2015 while the rest of the country dealt with a peak in late November 2015. The districts which experienced the outbreak early also tended to recover earlier on average, and by May of 2015 there were a negligible number of cases across the country. Based on this analysis, the research found that at its simplest, the change in enrolment between 2013 and 2015 does not follow a clear trend with the Ebola caseloads at the district level. This is shown in Figure 8 below.

The district with most Ebola cases reported was the Western Area Urban where enrolment across all levels decreased by 5.7%. However, Bonthe had reported just four cases in the data, while the enrolment across all levels decreased by 28.5%, the greatest reduction seen after the outbreak. Western Area Rural was in the top three for Ebola cases and their enrolment increased by 21.0%.

Figure 8. *Change in enrolment by district between 2013 and 2015 across all levels (bars), sorted by Ebola cases reported (line).*



This lack of a pattern holds at all levels and is similarly shown below in Figure 9 for the junior and senior secondary schools. Western Area Urban had the highest Ebola cases and some of the lowest growth, but across other districts cases and growth do not appear closely related.

Figure 9. *Change in enrolment by district between 2013 and 2015 at secondary levels (bars), sorted by Ebola cases reported (line).*

The lack of 2014 raw data prevented the research from going any deeper than district-level analysis, but it seems that the educational impacts of Ebola were not consistent across districts. That given, the hardest hit area, Western Area Urban, did see a noticeable fall in enrolment in the years following, while the Western Area Rural saw an increase. This may be explained by temporary migration out of the city itself during the epidemic, but without school-level data it is difficult to disentangle.

However, looking at the after-Ebola data of the next few years, some interesting patterns emerged. In the years following the reopening of the schools after Ebola, enrolment increased in the Western Area (with the exception of senior secondary in Western Area Rural, which fell). There was also persistently lower enrolment in primary schools in Port Loko and Kambia, two neighbouring districts in the north-west. Figure 10 shows that even a year after Ebola (2016) the shock of school closure could be felt in the enrolment trends in some of the districts as they struggled to get back to before-Ebola enrolment figures, especially at primary level. It was not until the launch of FQSE in 2018 that the country saw a full recovery from the epidemic, and enrolment levels rose higher than the 2013 totals at all levels as shown in Figure 11.

Figure 10. *Persistence of impacts — 2013 to 2016 changes.*

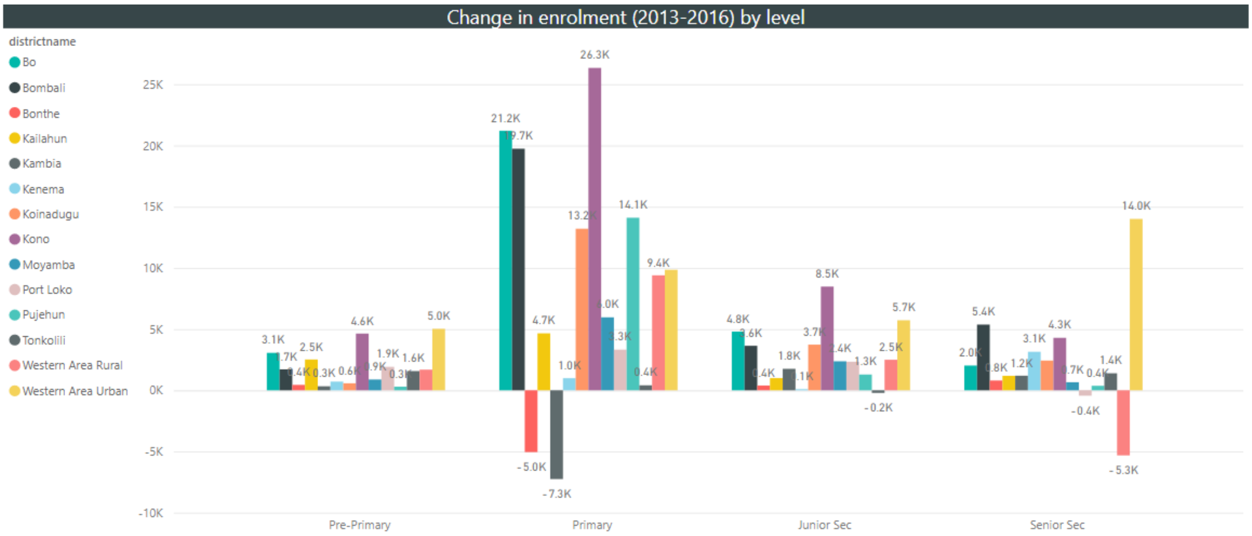
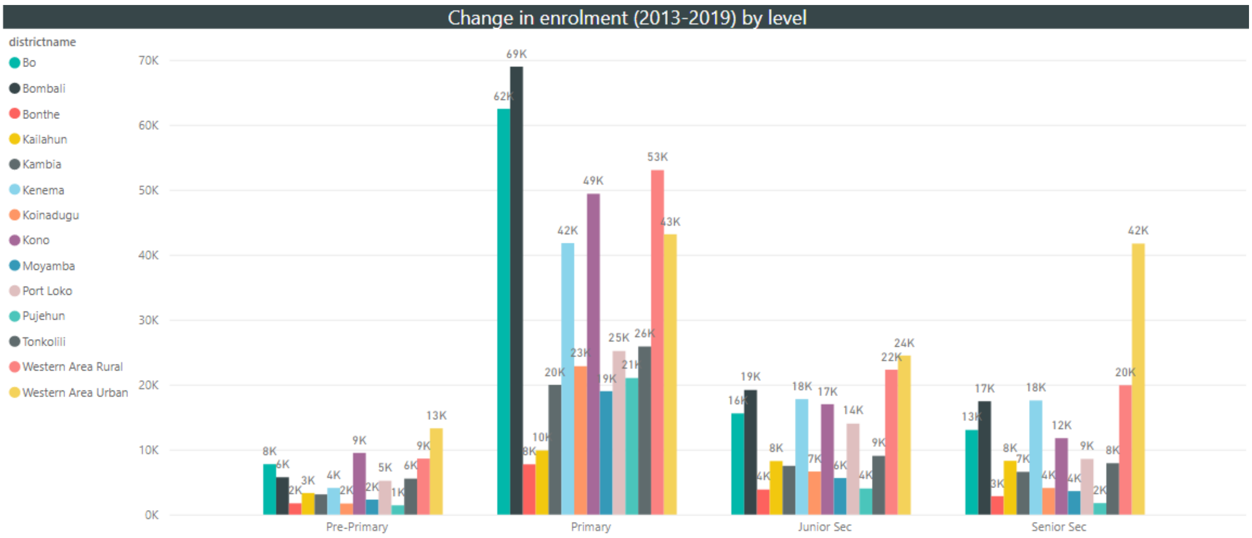


Figure 11. *2013–2019 enrolment changes.*



In finding any localised trends in enrolment before / after Ebola, the project also used spatial data analysis to look at spatial patterns in re-enrolment. However, spatial data analysis required detailed school-level data, which was only available for years after Ebola; the research, therefore, only looks at patterns post-Ebola, and how these evolved over time.

While the enrolment increased noticeably on average, driven largely by FQSE, surprisingly there were some areas where enrolment rates fell. This suggests high rates of transfers across schools, which may be due to parental expectations and abilities of schools to provide free education when not all schools receive government subsidies (looked at further below). The hotspot spatial data analysis could only look at how primary, junior secondary school (JSS), and senior secondary school (SSS) enrolment in schools grew over time. This analysis only included schools with data for all five rounds of the ASC. Figure 12 shows that the primary enrolment in schools across the country grew exponentially; however, looking at the higher levels, the positive enrolment changes become increasingly concentrated in the urban centres — notably the Western Area, Bo City, Kenema, and Makeni.

Figure 12. *Heatmaps of enrolment changes by level between 2015 and 2019.*

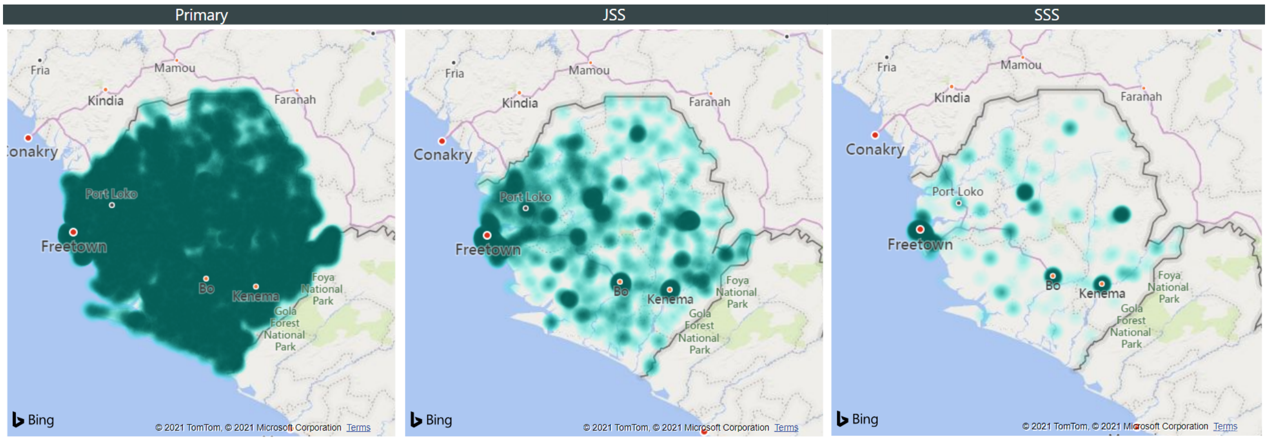
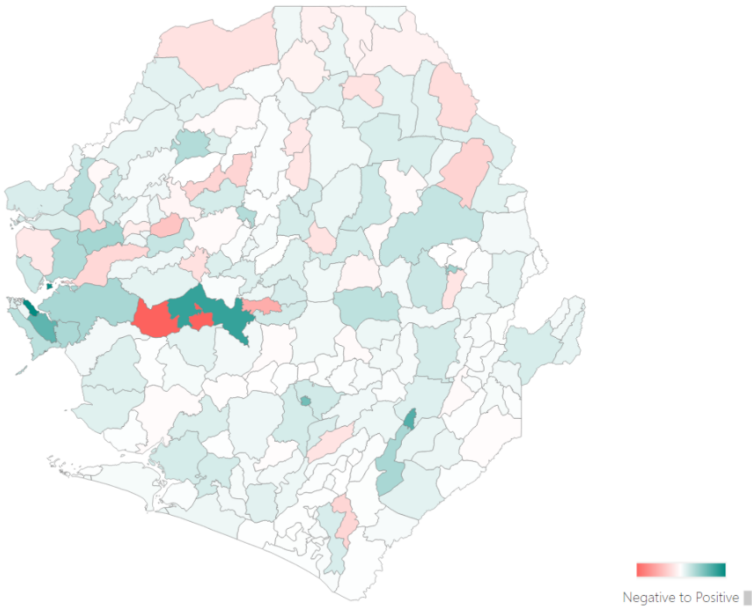


Figure 13. *Enrolment changes by chiefdom between 2015 and 2019.*



Alongside the absolute increases, the FQSE policy also saw a substantial movement of children across schools — away from unapproved non-financially supported schools into financially supported schools[9]. Looking at chiefdom-level changes over time (Figure 13), some areas were found losing children, while their neighbouring chiefdoms saw a noticeable increase. This is very pronounced in some chiefdoms in Tonkolili, where the enrolment fell, while their neighbours' enrolment increased by roughly the same amount.

[9] Not all approved schools receive full financial support and subsidies, but no unapproved schools receive that.

Table 5. *Proportion of approved schools and enrolment by year.*

Percentage of schools approved				
Status	**2016**	**2017**	**2018**	**2019**
Application made	33.25	32.54	11.03	19.30
Approved	42.77	43.69	51.19	66.85
Not approved	23.97	23.77	38.30	15.04
Total	100.00	100.00	100.00	100.00
Percentage of enrolment in approved schools				
Status	**2016**	**2017**	**2018**	**2019**
Application made	23.80	23.71	8.94	12.37
Approved	58.61	58.94	64.01	78.79
Not approved	17.59	17.35	27.04	8.85
Total	100.00	100.00	100.00	100.00

This movement is partially explained in the changes in enrolment trends in approved schools over the years. The enrolment in these approved schools increases noticeably over the period data are available (2016–2019), particularly after FQSE, when it rises to 79% of children being in approved schools compared to 59% in 2016 (Table 5).

The increase in enrolment is seen in all the districts[10] in Sierra Leone, both in absolute terms and in percentage changes. However, the magnitude of these changes varies, primarily driven by differences in population. It wasn't possible, within the time available, to merge in the high-density population data to calculate changes in gross and net enrolment rates over this period at the chiefdom and district level, but this seems a potential extension for the

[10] This analysis is complicated by the creation of two new districts in 2017. In each case this was the result of splitting an existing district into two. This has the result of making the original combined district look like it has decreased — but adding these back into their historical districts sees increases across all.

research that can be taken forward. In the third sub-question, possible extensions to this work are looked at to see how these past changes can be used alongside other available data to help identify areas of vulnerability.

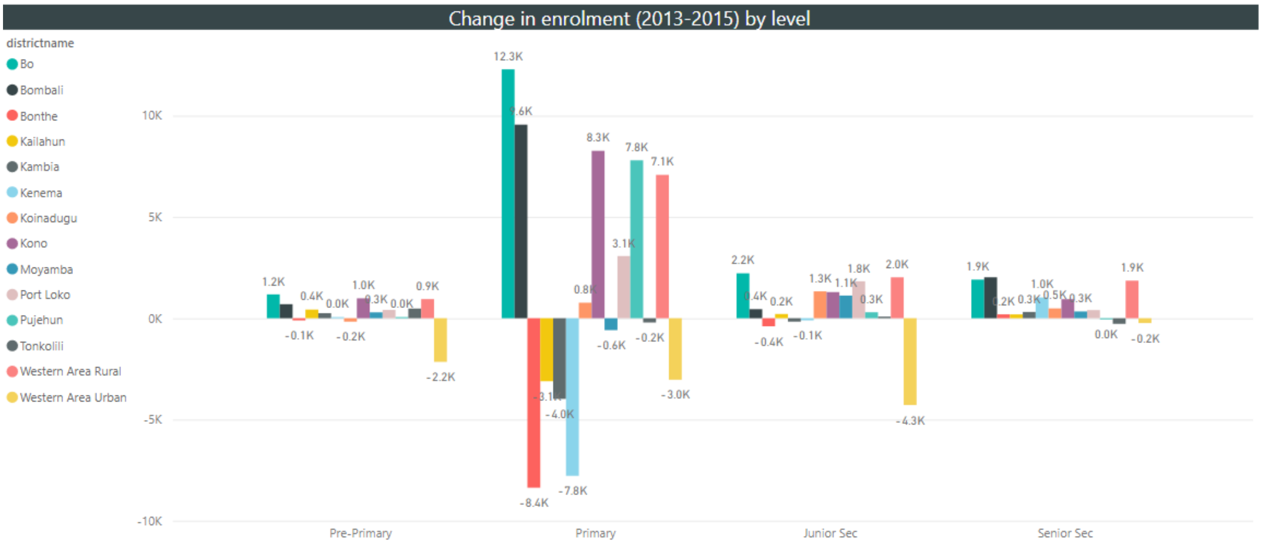
RQA3. Are there any differences by gender in terms of re-enrolment following Ebola?

In this sub-question the research looks at the gendered effects of enrolment changes and finds a greater increase in enrolment, on average, for girls at every level following the school closures than for boys. This is shown in Figure 14 below.

Figure 14. *Country-level absolute and percentage change in enrolment (2013–2015) by level and gender.*

However, again this hides local discrepancies, with girls' enrolment falling more than boys in a number of regions. This is shown in Figure 15. In the Western Area Urban girls' enrolment fell more than boys in primary, junior secondary, and senior secondary school. Interestingly, for junior secondary this is not reflected in an equivalent increase in Western Area Rural — and looking at these two districts together, the combined change in enrolment at junior secondary was -2,300 for girls and only -800 for boys.

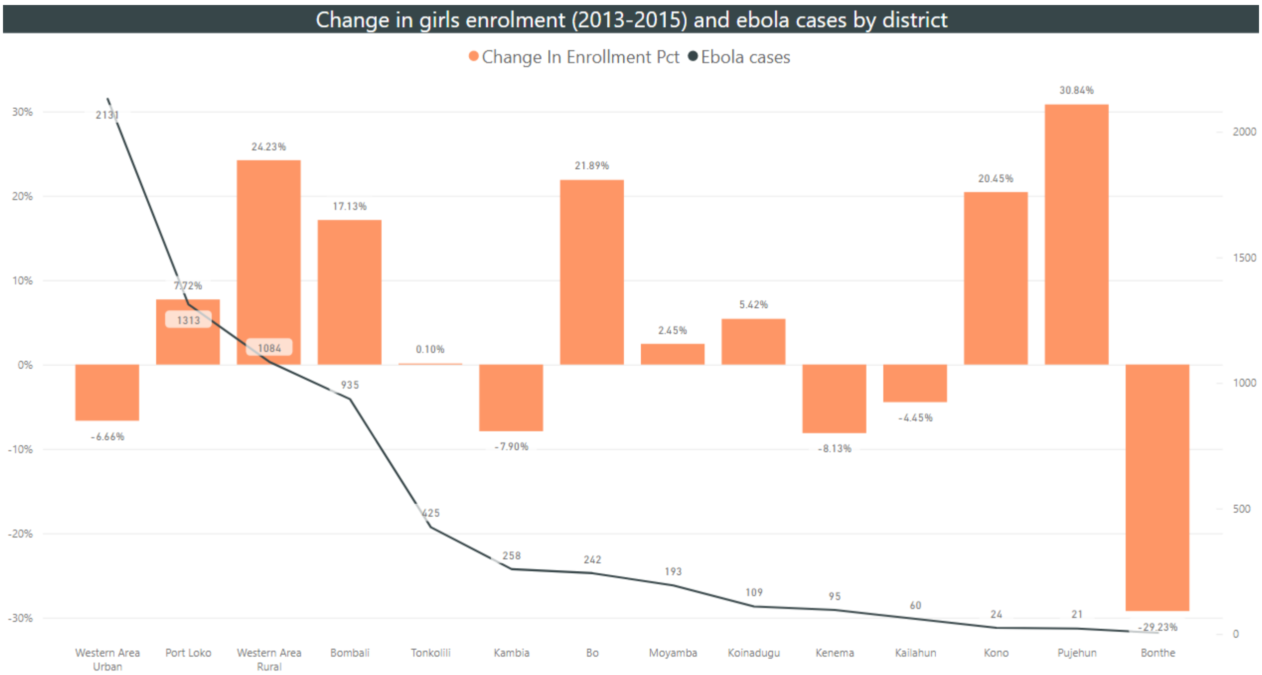
Figure 15. *District-level changes in enrolment (2013–2015) for girls by level.*



The existing literature found that many girls reported early pregnancy and marriage following the school closures. Unfortunately, this research cannot comment on this due to the absence of data on pregnancy and parenthood in the ASC[11].

Similar to the data for both sexes combined, the enrolment data for girls did not show any clear patterns associated with data on Ebola cases during 2013 and 2015. This is shown in Figure 16 below.

Figure 16. *Changes in girls' enrolment (2013–2015) and Ebola cases by district.*



[11] This has been newly included in the 2020 ASC.

RQA4. Is it possible to use these patterns, and household survey data, to predict likely areas of vulnerability going forward?

This sub-question required adding high-density information on relative wealth, produced by the Facebook Data for Good (D4G) initiative (↑Facebook, 2021). The D4G data was calculated for a three-mile catchment area of each school[12].

Inclusion of the relative wealth index in the database showed that pre-primary schools and senior secondary schools were usually to be found in the relatively wealthier areas with the relative wealth index increasing as the level of schooling increases, and with pre-primary schools being found in the highest relative wealth catchment areas.

Table 6. *Changes in relative wealth of financially supported schools 2018–2019.*

2018					
Financially supported?	**Pre-primary**	**Primary**	**Junior Sec**	**Senior Sec**	**Total**
False	0.23	-0.13	0.05	0.22	-0.02
True	0.39	-0.11	0.08	0.19	-0.05
2019					
Financially supported?	**Pre-primary**	**Primary**	**Junior Sec**	**Senior Sec**	**Total**
False	0.27	-0.15	0.10	0.23	-0.02
True	0.20	-0.08	0.06	0.23	-0.03

One way this data can be used is to look at the relative wealth of the catchments of the school receiving financial support compared to those that are not (excluding private schools) (see Table 6). Looking at this in the years between 2018 and 2019, it can be seen that the relative wealth of the catchment area of financially supported primary schools and JSS fell, suggesting the newly supported primary schools and JSS were in poorer

[12] To do this the GIS locations of the schools were used and the average RWI for the five-kilometre buffer for each school was taken.

areas. This suggests that the schools prioritised for financial support under FQSE were from poorer areas of the country. However, for SSS the relative wealth of the catchment area of a financially supported school increased slightly. This may suggest that given the limited number of SSS, the only schools that were not being previously supported by the government were in affluent areas, which under FQSE started receiving financial support as well.

The research also trialled Machine Learning (ML) approaches to see if the ASC data together with Multiple Indicator Cluster Surveys (MICS) data can be used to predict children that are at highest risk of dropping out of school. As detailed in the methodology, overage population in class and high grade-one population were used as two key criteria to identify the high-risk schools. Any school with more than 40% enrolment in grade one and having more than 25% of children overage in any grade was defined as high risk. Table 7 shows the schools that meet these criteria had an almost 12% decrease in enrolment year on year from 2015 to 2016, which, when compared to the overall primary school enrolment increase of 4.47%, shows that the overage and high grade-one enrolment characteristics can help identify schools where most at-risk children could be found.

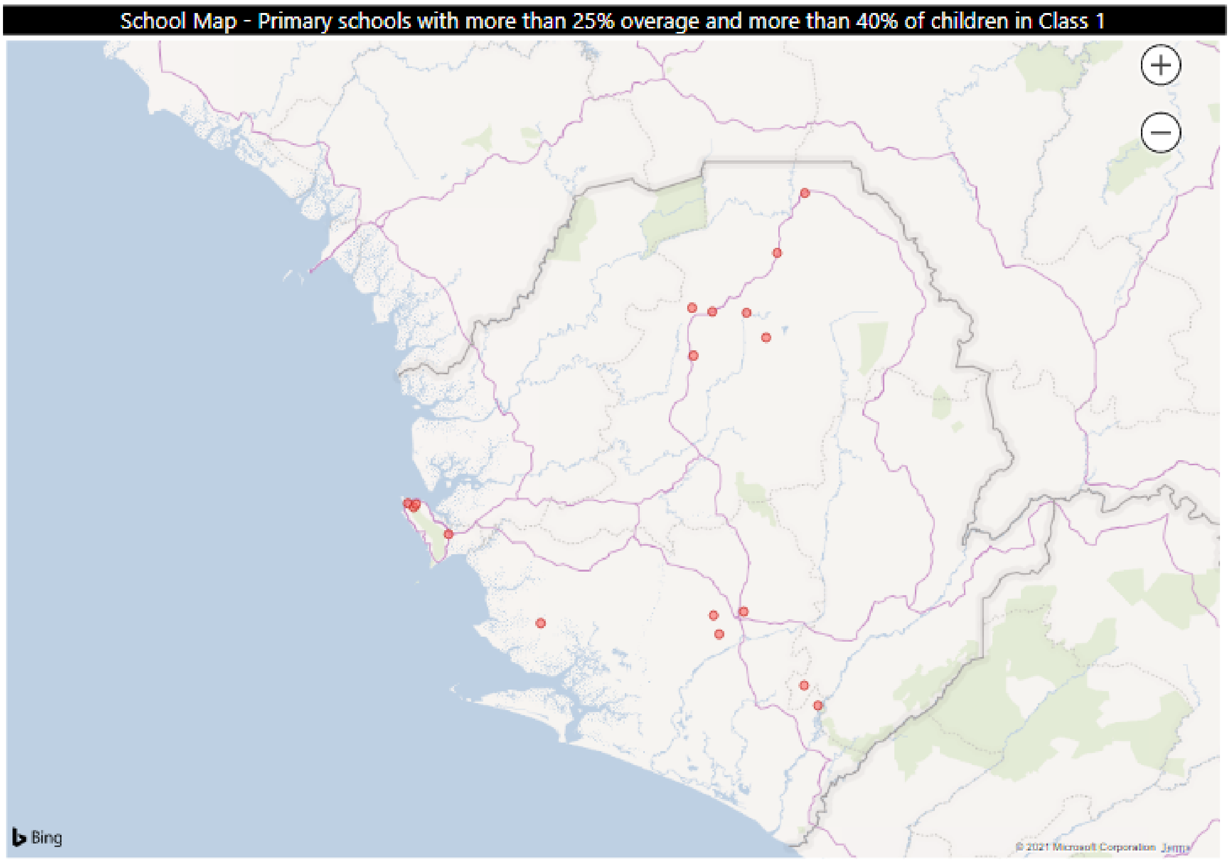
Table 7. *List of high-risk schools found using the ML.*

School name	Total enrolled	Class 1 enrolment	Class 1 share	Overage children	Overage share	Year-on-year % change in enrolment
National Islamic Primary	69	36	0.52	33	0.48	-22.47%
Islamic Welfare Union	61	27	0.44	21	0.34	-37.11%
Sierra Leone Muslim Brotherhood Primary School	129	56	0.43	34	0.26	-7.86%
Mentally Retarded Children Primary School, BO	43	20	0.47	19	0.44	-15.69%

Baptist Primary	142	58	0.41	38	0.27	20.34%
Ahmadiyya Muslim Primary	44	23	0.52	13	0.30	-58.10%
St George Roman Catholic Primary School	114	46	0.40	17	0.25	25.27%
Roman Catholic Primary School	67	28	0.42	17	0.25	-2.90%

Finally, the combined dataset was used to track the schools over time, so even though GPS data on schools was not collected until 2018, the location data could be backdated to allow for visualising this list of 2016 schools on the map (Figure 17).

Figure 17. *Schools with student populations which are at risk of dropping out.*



RQA1. What lessons can be drawn from the data on school-level changes in enrolment following Ebola, which can be applied to school reopening following Covid-19?

The lack of direct patterns between Ebola cases and school enrolment discussed in the sections above is in itself a lesson that can be applied to school reopening following Covid-19. Principally, it is not just the direct health impacts of epidemics and pandemics that are likely to affect enrolment, but a wider range of geographic and economic factors should also be considered.

For example, the greatest incidence of Ebola cases was in the Western Area around Freetown, with Western Area Urban having the highest number at 2,131 and Western Area Rural having the third highest number at 1,084 cases reported (with nearby Port Loko having the second highest). While the data does show a decrease in enrolment from Western Area Urban, this is exceeded by the increase in enrolment in Western Area Rural, suggesting there was a degree of transfer from the greatest population density in Western Area Urban to the surrounding areas instead of actual dropout.

On the other hand, despite having the lowest reported cases, Bonthe — one of the poorest districts across the Relative Wealth Index and other data[13] — was the district that had the greatest percentage fall in enrolment between 2013 and 2015 and was one of two districts not to have reached pre-Ebola enrolment numbers by 2016[14]. This suggests that the various barriers to education may not change greatly during health crises, and ongoing challenges, such as poverty and the economic impacts of the health crisis, might be of greater importance for education outcomes.

This is further supported by the greatest percentage increases in enrolment across levels between 2013 and 2015 being in pre-primary and SSS. In both these levels, only one district had a fall in enrolment[15] compared to seven and four districts in primary and JSS respectively. In pre-primary and SSS, there is a greater share of children from wealthier households, which appears to have cushioned the impact of Ebola on access.

In terms of gender, the data on enrolment finds a greater increase between 2015 and 2013 for girls than for boys across all levels. However, this does not mean that localised issues or particular challenges for girls during school

[13] The 'Sierra Leone 2015 Population and Housing Census Thematic Report on Poverty and Durables' from Statistics SL lists Bonthe District as having the highest multidimensional poverty rate (↑Statistics Sierra Leone, 2017).
[14] The other being Moyamba, also one of the poorest districts according to the Relative Wealth Index and Census.
[15] Western Area Urban for pre-primary, and Tonkolili for SSS, both of which were minor falls.

closures, such as a greater risk of early pregnancy found in other literature, should not be investigated and mitigated against going forwards.

Finally, although the school-level data was not available for pre-Ebola to enable more granular analysis of vulnerability in the context of school closures, the potential that geographic and economic factors may play a greater role in school access than direct health impacts during health crises suggests that geospatial analysis, dimensions of vulnerability, and their combination with ML could play a very useful role in mitigating education impacts in the future. These lessons are discussed further in the policy implication section below.

4.4 Second aim of the paper (B): RQB1 to RQB3

RQB1. What does the optimal education data architecture look like, and what are the key initial steps to take towards building such an architecture, given past experiences and the constraints on the field?

The starting point for defining any data system is to agree what information is needed, for what purpose, and how it will need to be presented. In education, there is general agreement on the core element of this, with Education Management Information Systems (EMIS) being prevalent in most countries (though they are of varying quality), with a relatively standard set of indicators, focusing mostly on enrolment, teachers, and school-level inputs. Increasingly systems are looking to incorporate learning outcomes into such systems but they are generally managed separately (both for exam pass rates and for sample surveys).

This section tries to outline how Sierra Leone, and other education systems, can move away from individual ad hoc datasets into a more structured database which allows for easy queries (to answer the questions of education stakeholders) and change tracking (to see how things are improving, or being implemented over time). It also discusses how to design a database from a technical standpoint and what key information the database needs to contain.

This section is partly informed by the research team's experience, partly through the literature, and partly through learning by doing — as a practical output of this work the team has created a Structured Query Language (SQL) database of the 2015–2019 data (and are integrating 2020 data). This has been accompanied by a data-analysis platform, currently hosted on Fab Data's SharePoint but with ongoing discussions for transferral and ownership by the government, which allows users to investigate the data more.

Mapping the education system

In order to plan and implement a successful and complete education data architecture it is important to map each of the components that will go into this data system, the data each component has available and the desired outputs of the system (i.e., school-level analysis or district comparisons), and to structure the system accordingly. As the number of possible components is very broad, this paper differentiates between those for education management for ministries and local governments, and additional optional components for specific departments / partners.

To develop an education data management system that keeps the record of all the schools in the country, there needs to be an agreement on all possible components and sub-components of a school. The list below provides some basic components as an example here.

- Schools.
  - School owners (Government / Private / Community / Missions)
  - School facilities
  - School finance
  - School staff — these link to the teacher records
- Staff
  - Teacher on payroll
  - Teacher qualifications
  - Teacher specialties
  - Non-teaching staff
- Students
  - Student enrolment
  - Student attendance
  - Student exam results / assessment results
  - Student demographics

These components will form the bases of an EMIS in the form of tables. Whatever the contents of these various tables (such as school, staff, students) the single most important data feature when implementing systems that are relational (or if any form of change tracking is required) is the need to have a unique identifier for each central record. The unique identifier can be in any format, but the key thing is that it **NEVER** changes. This is fundamental to the functioning of a Data Management System. In Education Data Management Systems, the unique identifiers could be:

- **Within education:**
  - School ID
  - Student ID (if you wish to track students over time)
  - Teacher ID
- **For administrative regions:**
  - Chiefdom code
  - District code
  - Local Council code

While these are the core areas, there are a myriad of other areas which can be included. For example, it is also useful to look at other areas, particularly the wider support network in education, which mainly comprises government staff (at central and local levels) and community support. For the government management system, including key staff (such as the inspectorate or school support officers) is valuable and allows for management of ancillary support services.

Outside of the components that are directly related to the education system, there is also a need for the data management system to include components from other ministries — some of these will be core components, such as the countries' administrative area zoning — districts, and chiefdoms; some will be optional but highly desirable, such as budgetary data and population data. At the local level this may include the education support staff (if education is devolved to a ministry of local government).

Designing the database

Once the system has been mapped, the next step is to think through the technical design of the database to be used to run the EMIS, which, if implemented, can underpin the smooth running of the education system. There are some fundamentals that any country developing an EMIS database should follow. These are detailed below.

*Data should be stored in a secure, well designed **relational** database. A relational database is required to ensure data integrity between records and to ensure the core data is entered continually.*

A relational database stores data in tables and rows (records) which are linked to each other via keys in each table — with child tables linked to their parents using the unique identifying column of the parent table. So, each school has a unique identifier, from which teachers can be linked from the teacher table (using their unique ID). For tables to be processed in a query, the database must be able to source the keys in both tables, enforcing data integrity. This is the most common form of database, with many options available, including open-source platforms such as MySQL. For this project Microsoft SQL server has been used to process the database[16] — alternatives include the OpenEMIS system, which is a low-cost initiative supported by UNESCO.

The alternative here is a non-relational database (such as NoSQL), which does not automatically require data to be stored as tables with associated keys. These have the advantage of being able to store huge amounts of data, and flexibility to change without having to redesign the system and links. The main advantage is that they can capture all data that are available, even if they are unstructured and unlinked but might become useful as information evolves. The disadvantage is that data can become orphaned easily (for example, teachers without schools or schools without districts), which is very high risk in operating environments where data entry is challenging and verification is costly.

Data from the database should be extracted by a collection of secure, flexible Application Programming Interfaces (APIs)[17].

Each API should have its own security requirements and be focused on specific functionality, for example, one API specific to school data, another specific to teacher data, and another specific to payroll.

The reason for this separation in the design allows for both focused upgradability — one part of the system can be upgraded without impacting others — as well as focused security. Ideally the access to the APIs can be configured so that only specific applications or users can access specific APIs and undertake actions. In the case of Sierra Leone, this will enable data specialists within institutions to have access to the relevant APIs and actions for their institutions, while policymakers and operational staff in decentralised offices have access to the outputs.

[16] This is not open source but was chosen for a number of reasons: familiarity with the software, flexibility in hosting either on a local server or in the cloud, and easy integration into tools such as Power BI.
[17] Application Programming Interfaces allow computer applications to talk to each other.

Focused applications should be built to interact with the data APIs to provide all of the required functionality for the users.

An example of this could be building separate web applications for each district education officer in Sierra Leone so that they have access to all the detailed data down to school level for their district, but only high-level aggregate reports for other districts for comparison. Such applications can take any form — they could be web applications, desktop applications, or mobile apps. The core detail here is that all applications would be talking to the same APIs, which in turn are *all talking to the same underlying data structures*. This means that although you have multiple user applications, they are all talking to the same underlying data providing a single source of truth for all of them (a principle discussed in more detail in Section 4.4, RQB3).

Alongside the relational database, there should be a data warehouse that would be designed to poll the relational database daily.

This data warehouse would be designed to be optimised for use in business analytics services and would handle the change tracking where it was required. This data warehouse method to showcase the data is currently being explored by the Ministry in collaboration with other actors in Sierra Leone.

Similarly, Sierra Leone (among other countries) is increasingly working towards real-time data collection that is updated regularly, instead of only annual data collection efforts. In this case, the data warehouse would query the relational database each day to pick up new or changed records.

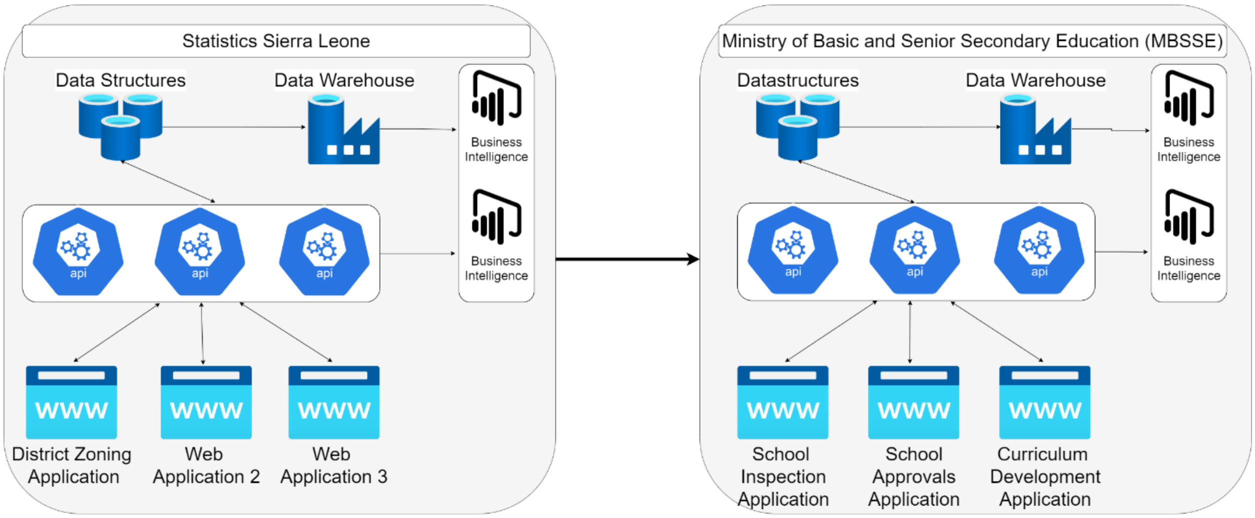
Lastly there would be a collection of data visualisation / querying solutions that extract and present curated versions of the data warehouse data to end users.

This final step is key to usability and should be structured so as not to expose the complete data warehouse but as curated datasets to satisfy specific user needs. Here there are a number of options, with the main choice being between coding a bespoke dashboard and utilising Software as a Service (SaaS) providers of business analytics software.

The former has the advantage of not requiring licences, which on the surface means that it can be cheaper. However, the cost of licences should be traded off against the cost of development and adaptation for users' needs. For example, a licence for Power BI costs approximately £10 per month for private reports (public reports have no requirements for licences to access), while a software developer to manage a security login system and recode data visuals can cost many multiples of that per day. If the software developer is needed for two days per month, then even a hundred users could access a highly customisable package for the same cost.

Figure 18 below shows these principles visually[18], focusing on how the education sector data can be structured to sit alongside the wider data systems in Sierra Leone.

Figure 18. *Potential optimal data architecture for the education system in Sierra Leone.*



Within this architecture, each department has their own collection of data structures, services, and applications. Importantly, as shown by the connecting arrow between the departments, data only flows in one direction when crossing departments. What it means is that data can only be read from another department, and it does not have the ability to make changes. In the case of location data, an application within the MBSSE can request a location, for example using GPS data or a town name, and receive back the current chiefdom and district as defined by Statistics SL. It could not, however, change where a town or GPS location is defined to be. Another benefit of having this system / setup would be to prevent data errors. As seen throughout ASC datasets, schools repeatedly give incorrect location data regarding the chiefdom and district combinations. This requires cleaning of the data every single year.

Designing the processes for data collection, entry, and validation

Another set of considerations (on top of the data architecture design) are the processes for data collection, entry, and validation that can be refined. One key area is how new schools are added to the system. Prior to 2018, all of the schools in the school census were allocated unique, non-duplicated EMIS codes. However, from 2018, a significant number of schools were no longer assigned EMIS codes, and the schools were given generic codes upon data

[18] This diagram is not 100% complete as other items would need to be taken into account, such as file storage. Moreover, the number of icons for applications, APIs and Business Intelligence are only illustrative.

entry. As highlighted before, without the unique IDs the system breaks, requiring lots of effort to match schools manually.

Practically this means that for the data system to work well, the identifiers should not just be used for the annual census but integrated into other areas, such as the school approvals policies and processes, and catchment planning; so, if a new school is created by a community and it seeks approval, then it should be immediately assigned a unique ID. A similar process can be conducted for teachers, who should be assigned an ID when they register with TSC.

Finally, how school data is collected and managed requires more thought. Currently the school census is collected every year by a team that goes into the field. It means updates to school information only occur once a year. A more optimal solution would be to have the schools updating their own information via a tablet, something which is being discussed within the One Tablet Per School (OTPS) initiative and an area which increases the possibilities for the data system exponentially.

RQB2. How can we demonstrate how the Annual School Census (ASC), built with a well-planned new data architecture, can help the government in improving education planning by using existing ASC data?

In order to demonstrate the potential of a well-planned data architecture and more consistent data collection processes, the project created a functioning SQL database merging the 2015 to 2019 data from its previously disparate sources. This database will be transferred to the Directorate of Planning and Policy at MBSSE, and can form the basis of future work and analysis. This database has been in high demand and has been used already to inform decision-making and will be used going forward by the Ministry, TSC and development partners. It has been used to create data-for-communications products, which have been shared with the Minister, who has approved the publication of the database on the Education Data Hub portal and the integration of the 2020 data.

In this section, the steps taken, the final database output, and the front-end that was created to showcase the potential of the database are each described in detail.

Combined ASC 2015–2019

A very useful output of this study is the compilation of the 2015–2019 ASC datasets into one combined database, which is securely stored in an SQL database and will be transferred to the MBSSE once a secure storage solution is agreed.

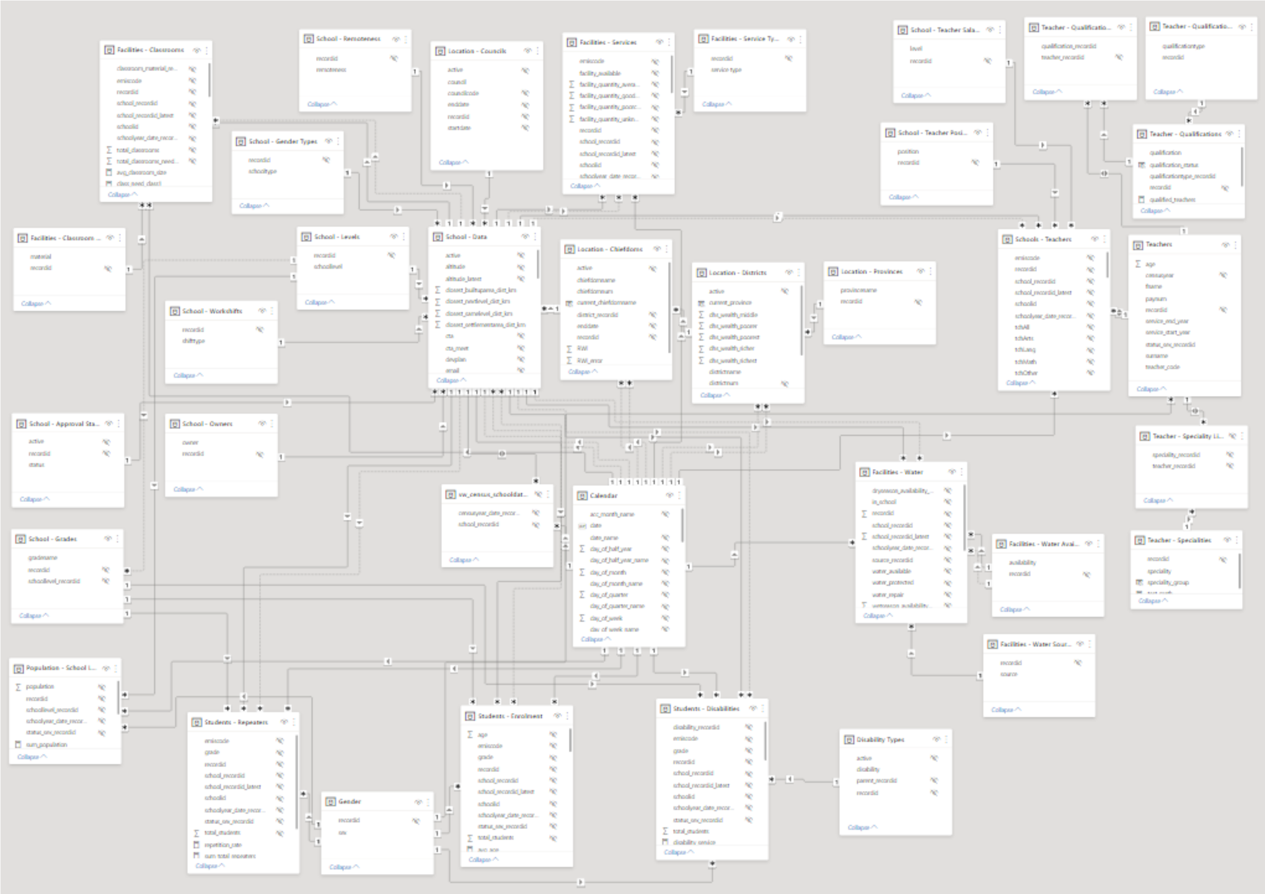
To maximise the efficiency of storage and the speed of analysis, the data is organised in a database structured around schools, which are matched over time. To recall, SQL deals with relational databases, where individual data-tables are stored separately and then linked using 'joins' to enable data processing and extractions, which have been defined and structured to enable rapid queries.

As discussed in the optimal data architecture, for education data the optimum central table is the school table, which joins to the other tables (such as the teacher table, or the enrolment table). These tables can in turn be linked to other information. For example, the enrolment table can be linked to the disability information, and the teachers table can be linked to the teacher qualifications table.

This might seem minor, but having each category of information stored in separate tables makes life easier for querying (as less information needs to be transferred) and managing the storage size (as less information needs to be saved). For example, in a relational database there is no need to save the school name multiple times across each table as each table is linked based on a school number (which takes less space to save) and this number is linked to a name for when it is needed. These small changes can make processing substantially quicker, which is key in resource constrained environments.

The SQL database structure follows this logic, and the relationships between the tables is visualised in Figure 19 below (a high-resolution graphic of this schema can be downloaded via this link for better readability). This infographic is intended to illustrate rather than to show the full breadth of detail. Nevertheless, it highlights the number of relationships that are needed for an education database with this breadth of information collected with varying levels of aggregation.

Figure 19. *Sierra Leone combined education database (2015–2019) schema.*



Having this database available will dramatically improve the speed and reliability of data analysis in Sierra Leone, which can underpin a range of policy and programming areas — this database has been very well received and is already being used in a variety of settings.

Building on this is a linked output which wasn't in the original workplan: an interactive dashboard where the data is curated and presented simply. This was built using Power BI, which allows quick compilation and curation of the database at a lower cost than hard coding the visuals. The database is based on the combined database above, with the data curated to provide the information as clearly and visually as possible to maximise the use and understanding by the end users including MBSSE and TSC national-, district-, and even school-level staff. This is a by-product of the work the team was undertaking to conduct the research, with the additional effort of curating the dashboard being very small compared to the effort to compile and structure the data. As the team are working in Sierra Leone on a series of linked issues, this curation has enabled direct integration of the research findings into discussions and workflows.

The interactive dashboard allows users to easily look into more detail of the specific districts, genders, and school types (including by level, owner, and approval status) that they are most interested in, and to look at differences in trends between these. This is best experienced by using the dashboard directly, but screenshots are presented here. On each page, a wide range of tabs can be clicked on to see different aspects of that data.

Figure 20. *Homepage of the interactive dashboard.*

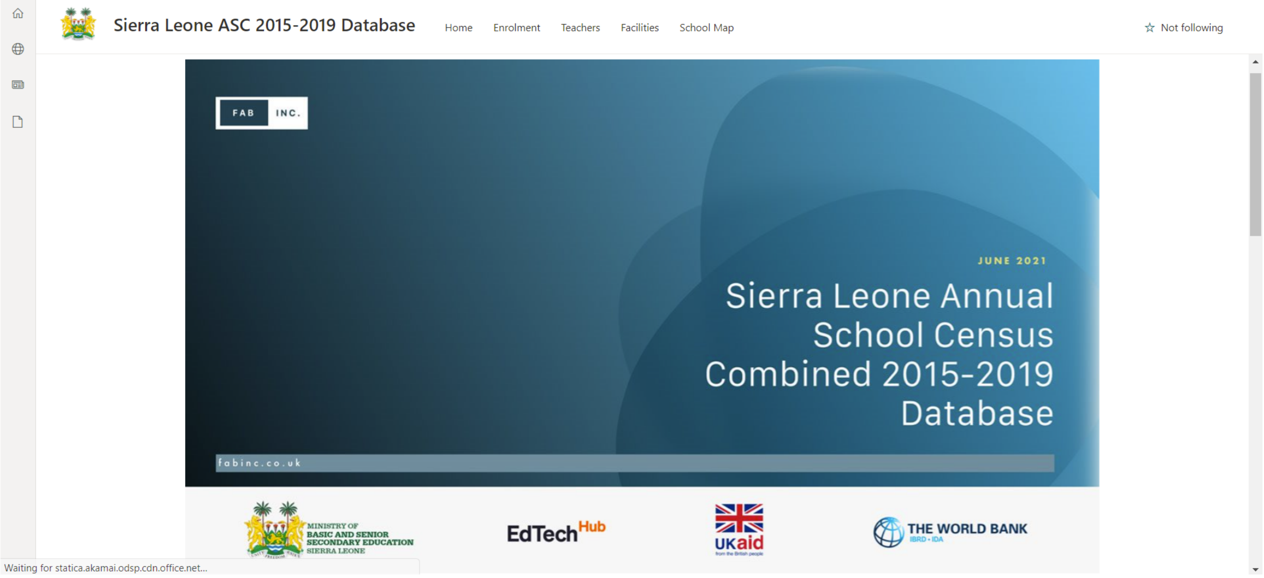


Figure 21. *Homepage of the interactive dashboard.*

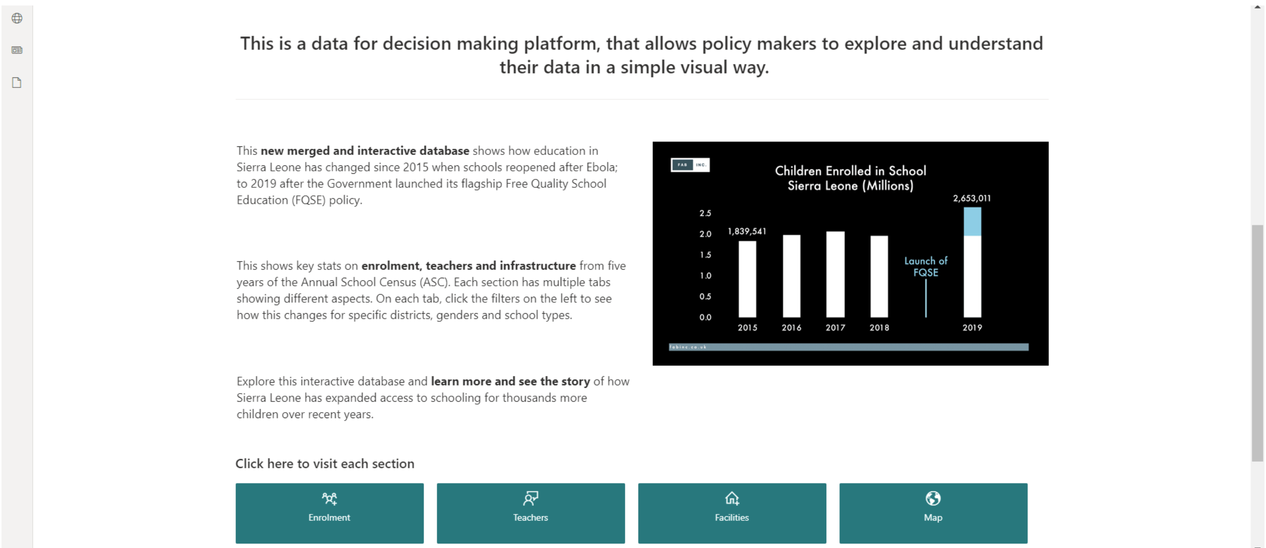


Figure 22. *Enrolment page of the interactive dashboard.*

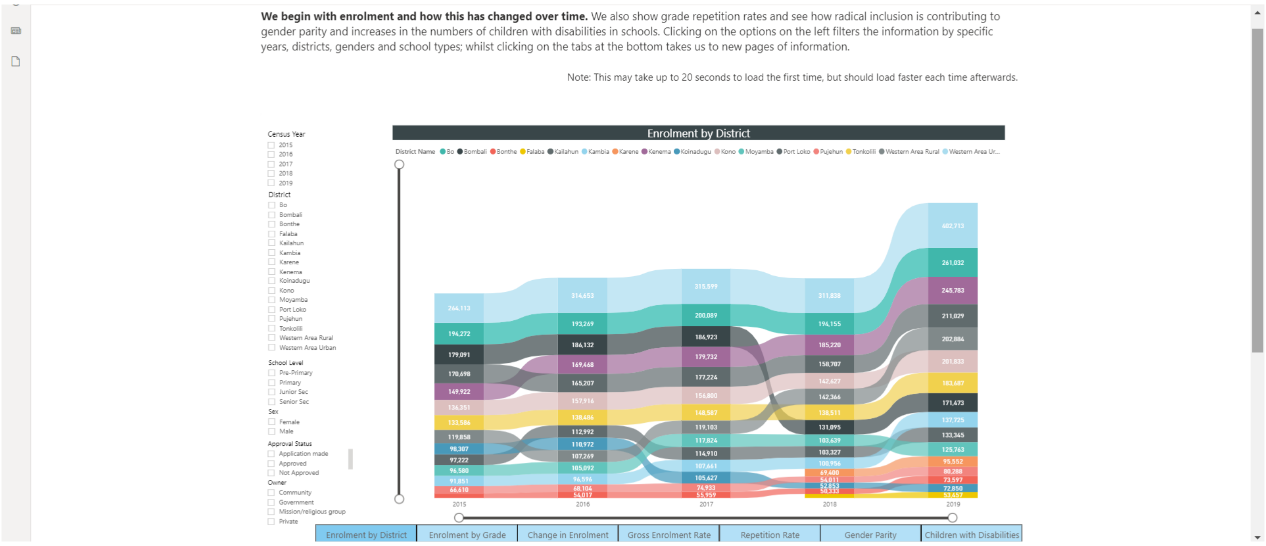


Figure 23. *Teacher page of the interactive dashboard.*

Figure 24. *Facilities page of the interactive dashboard.*

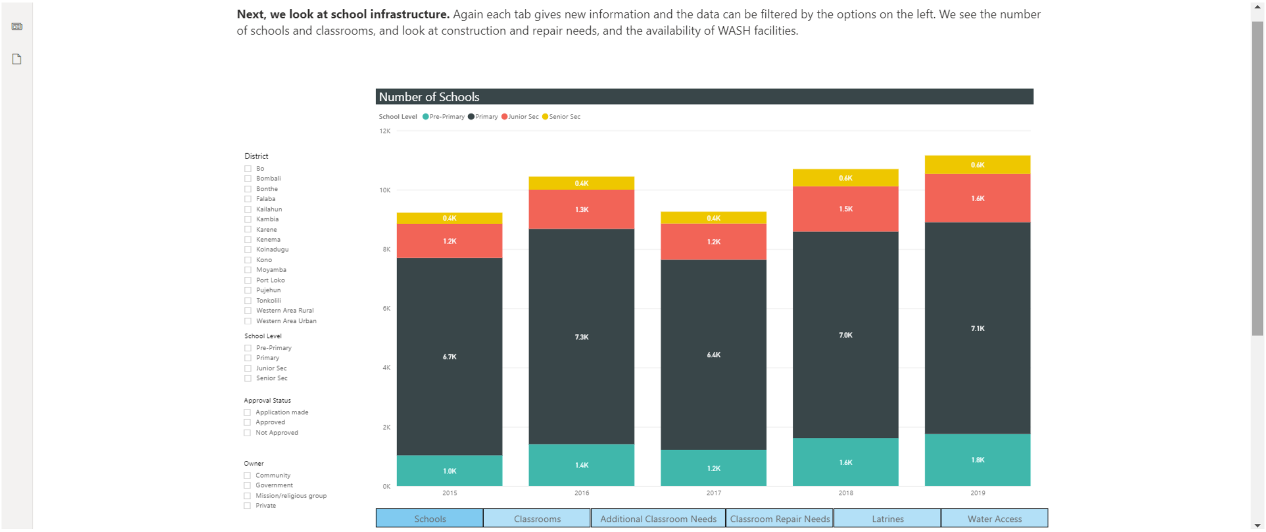
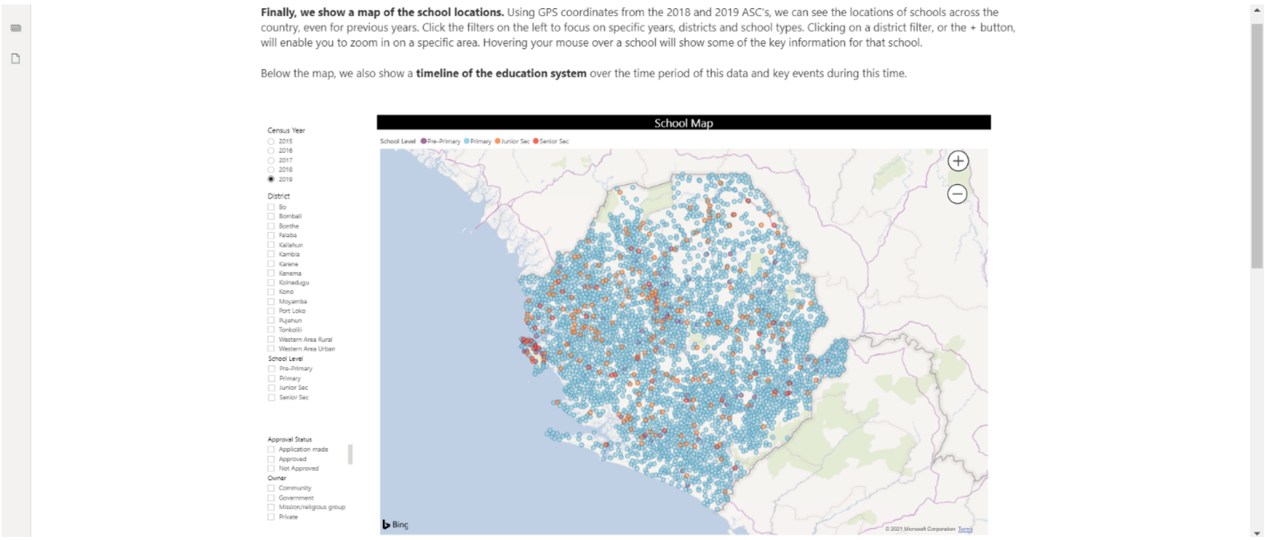


Figure 25. *School map page of the interactive dashboard.*



The dashboard is intended to be interactive for the data users, including government staff and other stakeholders in Sierra Leone, to enable them to examine and interrogate the data themselves.

The team is discussing with government staff how to transfer the hosting of both the database and the dashboard, to enable ongoing updates and amendments.

RQB3. What are the key components / datasets for this system — which currently exist, what are planned, and how can they be linked?

The current context of data systems in Sierra Leone is complex. There is a proliferation of systems which are often fulfilling multiple aspects of the data architecture for different users, but crucially are not interacting between these systems. This is therefore resulting in an overall data architecture for the education system that is not fulfilling the two key principles of data management. These principles are:

- Don't Repeat Yourself (DRY) — Every piece of knowledge must have a single, unambiguous, authoritative representation within a system. Any repetition can unnecessarily increase the memory needed for storage, and more importantly could also allow the possibility of contradictory information about the same item and the violation of the Single Source Of Truth (SSOT) principle.
- Single Source Of Truth (SSOT) — Every data element is mastered or edited in only one place. Any possible linkages to this data element are by reference only and therefore all refer back to the single source of truth.

This current context has a number of ongoing, piloted, and planned systems. As part of this research the following systems were identified:

- Annual School Census (ASC) — ongoing
- Teacher Management Information System (TMIS) — planned
- Payroll systems — ongoing
- Education Data Hub — ongoing
- Sierra Leone Education Attendance Monitoring System (SLEAMS) — piloted
- Teacher Record Management (TRM) system — piloted
- Leh Wi Lan Tangerine — ongoing
- Vatebra Joint Education Sector Dashboard — planned
- This 2015–2019 ASC Database — ongoing

And the key institutions and stakeholders involved in the development and use of these systems include:

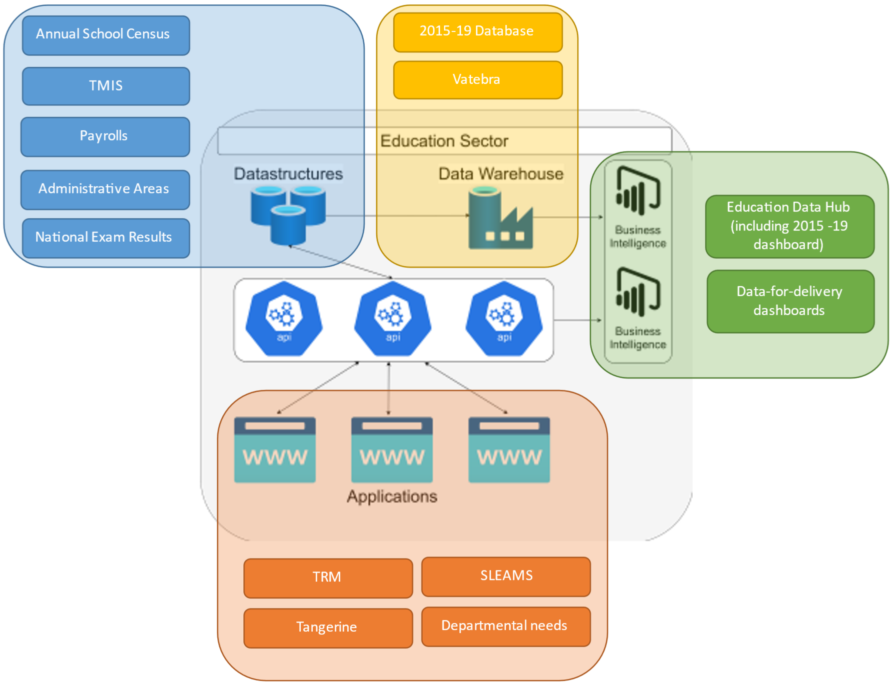
- MBSSE, TSC, MTHE, DSTI, Ministry of Finance, Statistics SL, development partners and projects

Another important initiative in this area includes the OTPS initiative piloted and planned for further roll out in coming years.

A key constraint within education data management in Sierra Leone is the self-contained nature of these systems, which are often fulfilling multiple roles within the data architecture but in an incomplete fashion and without interacting with each other to fill their gaps. Taking steps towards a more optimal data architecture will likely require greater specialisation and interaction of systems. Mapping the current, piloted, and planned systems to that architecture can help to identify potential intermediate steps and identify areas for further development.

Figure 26 below illustrates how the above-mentioned system might be integrated together as a single data system for the whole education sector in Sierra Leone. However, it would be useful for practical purposes that some of these systems are housed separately and only input data to the EMIS.

Figure 26. *Potential education data ecosystem for Sierra Leone.*



5. Policy implications

This section is divided into two distinct sets of policy implications, the first focusing on enrolment trends and impact of Ebola, and the second focusing on the use and management of the data architecture.

Based on the analysis of 2013 to 2015 enrolment data and the longer-term 2016 to 2019 data, it is evident that the effects of long-term school closure are localised, with poorest and remotest areas being the most affected. Therefore, it is necessary for any analysis to look at the localised trends in order to understand how the crisis affects the wealth and education of the households in different areas. Such localised analysis can then help governments develop localised solutions to ensure that the most vulnerable are prioritised.

Although it wasn't possible for this analysis, it would be worthwhile to understand the economic impact of these public health crises and how they affect the schooling of the children in different areas of the country. In the case of Sierra Leone, districts like Bonthe recorded fewer than 60 Ebola cases in total; however, the enrolment dropped by over 15,000 students across all the levels. This may suggest that the economic impact of the endemic forced the families to either move out of the district or take the children out of schools due to economic pressures, whereas in relatively wealthier districts of Western Area Urban, which recorded more than a thousand Ebola cases, enrolment continued an upwards trend between 2013 and 2016.

Similarly, the effects of Ebola on enrolment were also seen to vary by gender, with boys being adversely affected compared to girls. Though it is not clear what may have caused this, it could be that more boys were forced to leave the school and join the labour force to support their families financially compared to girls.

Sierra Leone's FQSE initiative has undoubtedly been a success in overcoming the negative effects of Ebola on enrolment, and the data from the next school year (2020 but pre-Covid-19) suggests these positive impacts of the FQSE policy were maintained in the medium term. Yet this policy has created substantial pressures, particularly in terms of the classroom– and pupil–teacher ratios, on already under-resourced schools. While this increase in enrolment has been matched with increased expenditure, there is a real need to bring in more qualified teachers onto the payroll and train existing teachers that are not qualified, and to ensure that classroom sizes in the early grades are more manageable. This analysis did not look at how these sudden increases in enrolment affected the learning of the students in classrooms. It's also worth noting that free schooling policies also triggered student movement between schools and at times between chiefdoms and districts, with students leaving fee-charging schools for government-subsidised free schools. This means that any policy changes which affect the desirability of individual schools should be modelled out, as it can create substantial pressure on schools who face an influx of students.

Policymakers in Sierra Leone, given their ambitious education reform agenda, require access to quality and timely data to ensure that the resources are used to reach the most vulnerable. Thus, the country can no longer afford to have poor quality, fragmented data that only provides the policymakers with a partial view of the system. A well thought out EMIS and data architecture will therefore serve as a foundation of Sierra Leone's education system.

Sierra Leone's timely investment in planning tools like the consolidated longitudinal database and interactive dashboard can help them make education planning and programming effective and efficient. This database, along with the dashboard, can help the Ministry of Basic and Senior Secondary Education (MBSSE) and the Teaching Service Commission (TSC) in identifying schools with the highest need for qualified teachers and subject specialists and help the government in deploying them effectively. Similarly, in a resource-constrained environment like Sierra Leone, such a data system can help the government experiment with innovative policies such as the employment of circuit teachers — specialist and highly qualified teachers who are strategically placed to meet the need of multiple schools in close proximity.

The Government of Sierra Leone already acknowledges that data is a key input to implementing its ambitious reform agenda, including the new policies such as the Radical Inclusion Policy, School Approvals Policy Guidelines, and the School Infrastructure and Catchment Area Planning Policy. Focus on the political and technical challenges of the data architecture and effective data management will maximise the impact of each of these policies and better enable Sierra Leone to meet the challenges of Covid-19 and future shocks.

5.1 Recommendations and next steps for the data architecture

Sierra Leone has already started a number of initiatives to ensure reliable and timely data is available to the policymakers. Such initiatives include the fully integrated 2015–2020 ASC database, the Vatebra Joint Education Sector Dashboard, One Tablet Per School (OTPS) and TSC-led Teacher Management Information System (TMIS). However, one area that requires further development is the development of a data warehouse. Neither the 2015–2020 ASC database nor the Vatebra Joint Education Sector Dashboard in their current scope would meet the full needs of the data warehouse, but the

2015–2020 database can serve as a solid foundation to build a data warehouse from.

Another key aspect of any successful data architecture is interaction and integration. As the institution with the broadest mandate within the education sector, the Directorate of Planning and Policy within MBSSE is best placed to manage the central education data system and provide access to required data to all relevant directorates within the education ministry as well as other ministries where required. Similarly, the Directorate of Planning and Policy should also work towards ensuring relevant data from other ministries is fed into the education data system on a regular basis. Recent progress has been made in this respect, with MBSSE and TSC collaborating well on the roll-out of OTPS and the planning of a Data Seminar incorporating stakeholders from across the education sector. One of the initial priorities for this collaboration could be on clearly setting out the data structure[19] requirements of each institution, for others to follow as an initial step on the road to greater integration. Moreover, OTPS can be a key catalyst for change, given the broad potential for changing data collection habits that it brings, and the breadth of uses available for bringing in and engaging the various stakeholders in formalised and interactive processes.

In the longer-term, the data structures for each entity (schools, teachers, students, exam results, payrolls, locations) should be defined by the institution directly accountable (for example, students are defined by teacher, teachers are defined by TSC or Head Teacher, schools are defined by DEO and Head Teacher). These data structures would then define what different data systems (composed of these data structures) would look like. For example, a system that only has information on teachers and students would be defined by the information gathered on the students, the information gathered on teachers, and the information gathered on the relationship between students and teachers.

Finally, there is very little structure to how the school census data is currently being collected in Sierra Leone. On top of this, the data collection procedures and scope vary greatly year by year. Once the handover of the 2015–2020 ASC database to the government has been agreed, the combined database can be transformed into a more structured system, in which users agree a specific format for the data to facilitate smooth import into this database in future. This structured format has the advantage that any third party could then collect the future ASC data however they see fit, but they would be required to deliver it in the specific format that can be imported into the existing

[19] Here 'data structure' means what kind of information needs to be gathered for different entities within the education system, such as school, students, teachers, DEO, CEO and how these entities are related to each other.

integrated ASC database. This would allow the government staff (Directorate of Planning and Policy) to run standard import routines (these will have to be written once the final data format has been agreed) year by year, which can greatly streamline data ingestion.

It is also advised that, as part of the discussions on handover of the 2015–2020 ASC database, the team also works with the MBSSE to agree the core and secondary components of the database going forward, and how this links to the school census data forms — alongside the data architecture work, the sheer volume of information collected makes it unwieldy. The system can benefit from the core database becoming more focused and defined.

6. Conclusion

This report adds to the learning from both the experience of Ebola and data management efforts to help inform post-Covid-19 recovery, including a data architecture for a resilient education data ecosystem in Sierra Leone and elsewhere. This was built on the construction of a combined database containing Annual School Census (ASC) data at the school level from 2015–2019, and at the district level from 2011–2013.

This combined database has enabled analysis of enrolment changes and Ebola cases across the country around the school closures of the Ebola epidemic, the gendered aspects of these changes, and the potential of using data to predict vulnerability, in order to consider lessons for the post-Covid-19 recovery in Sierra Leone and more broadly. This analysis highlighted the ongoing importance of geographical and economic factors in determining education access, over and above the short-term health impacts of such crises.

This work has also enabled a detailed consideration of the data architecture within Sierra Leone to inform ongoing stakeholder discussions in the education sector, as well as demonstrating the potential benefits of improved data structures for data analysis, visualisation, and use through a dashboard. This has been warmly received, with Minister David Moinina Sengeh presenting some of the findings of the analysis at the Global Education Summit, and the government looking to take over ownership of the database and intending to integrate the dashboard on their public Education Data Hub and in data-driven processes more widely. In doing so, the government has opened the door for EdTech Hub to support rationalisation and consolidation of further education data in Sierra Leone.

Finally, these findings are also likely to be relevant to many other countries which face similar challenges to Sierra Leone, both in terms of data availability and consistency, and Covid-19 school reopening. Key principles and step-by-step guides for replicating this work in other contexts have been provided to maximise the use of this research and support similar efforts in other contexts.

Bibliography

Abdul-Hamid, H. (2014). *What matters most for Education Management Information Systems: A framework paper* (No. 7; SABER Working Paper Series). World Bank. https://openknowledge.worldbank.org/bitstream/handle/10986/21586/944490WP00PUBL0Framework0SABER0EMIS.pdf?sequence=1&isAllowed=y

ACAPS. (2016). *Beyond a public health emergency: Potential secondary humanitarian impacts of a large-scale Ebola outbreak*. ACAPS. https://www.acaps.org/sites/acaps/files/products/files/t_beyond_a_public_health_emergency_feb_2016.pdf

Bandiera, O., Buehren, N., Goldstein, M. P., Rasul, I., & Smurra, A. (2019). *The economic lives of young women in the time of Ebola: Lessons from an empowerment program*. World Bank. https://doi.org/10.1596/1813-9450-8760

Cassidy, T. (2006). *Education Management Information Systems (EMIS) in Latin America and the Caribbean: Lessons and challenges*. VIII Regional Policy Dialogue Meeting, Education Network, Washington, D.C. https://publications.iadb.org/en/publication/education-management-information-systems-emis-latin-america-and-caribbean-lessons-and

Ellison, R. (2004). *A practical guide to working with Education Management Information Systems: Guidance for DFID advisers* (first draft). http://www.paris21.org/sites/default/files/2402.pdf

Facebook. (2021). *The relative wealth index*. Facebook Data for Good. https://dataforgood.fb.com/tools/relative-wealth-index/

Hallgarten, J. (2020). *Evidence on efforts to mitigate the negative educational impact of past disease outbreaks* (No. 793; K4D Helpdesk Report). Education Development Trust. https://opendocs.ids.ac.uk/opendocs/bitstream/handle/20.500.12413/15202/793_mitigating_education_effects_of_disease_outbreaks.pdf?sequence=6

Hua, H., & Herstein, J. (2003). *Education management information system (EMIS): Integrated data and information systems and their implications in educational management*. [Conference presentation] Annual Conference of Comparative and International Education Society, New Orleans, LA. https://www.infodev.org/infodev-files/resource/InfodevDocuments_188.pdf

Lassibille, G., Tan, J.-P., Jesse, C., & Van Nguyen, T. (2010). Managing for results in primary education in Madagascar: Evaluating the impact of selected workflow interventions. *The World Bank Economic Review*, 24(2), 303–329. https://doi.org/10/b4x2w7

Malala Fund. (2020). *Girls' education and Covid-19: What past shocks can teach us about mitigating the impact of pandemics*. Malala Fund. https://downloads.ctfassets.net/0oan5gk9rgbh/6TMYLYAcUpjhQpXLDgmdIa/3e1c12d8d827985ef2b4e815a3a6da1f/COVID19_GirlsEducation_corrected_071420.pdf

MBSSE. (2018). *2018 Annual Schools Census Report And Statistical Abstract*. Ministry of Basic and Senior Secondary Education. https://mbsse.gov.sl/wp-content/uploads/2020/03/2018-Annual-Schools-Census-Report.pdf

Murimba, S. (2005). The impact of the Southern and Eastern Africa Consortium for Monitoring Educational Quality (SACMEQ). *Prospects*, 35(1), 91–108. https://doi.org/10.1007/s11125-005-6822-z

Nayyar-Stone, R. (2014). Using national Education Management Information Systems to make local service improvements: The case of Pakistan (No. 30; PREM Notes). World Bank. https://openknowledge.worldbank.org/handle/10986/18416

Powell, M. (2006). *Rethinking Education Management Information Systems: Lessons from and options for less developed countries* (No. 6; InfoDev Working Papers). InfoDev. https://www.infodev.org/infodev-files/resource/InfodevDocuments_504.pdf

Santos, R., & Novelli, M. (2017). *The effect of the Ebola crisis on the education system's contribution to post-conflict sustainable peacebuilding in Liberia*. University of Sussex. https://educationanddevelopment.files.wordpress.com/2016/06/liberia-report_march2017_lowres.pdf

Smith, W. C. (2021). Consequences of school closure on access to education: Lessons from the 2013–2016 Ebola pandemic. *International Review of Education*, 67(1), 53–78. https://doi.org/10.1007/s11159-021-09900-2

Statistics Sierra Leone. (2017a). *Sierra Leone Multiple Indicator Cluster Survey 2017, survey findings report*. Statistics Sierra Leone. https://www.statistics.sl/images/StatisticsSL/Documents/sierra_leone_mics6_2017_report.pdf

Statistics Sierra Leone. (2017b). *Sierra Leone 2015 population and housing census: Thematic report on poverty and durables*. Statistics Sierra Leone.

https://www.statistics.sl/images/StatisticsSL/Documents/Census/2015/sl_2015_phc_thematic_report_on_poverty_and_durables.pdf

Sultana, R. G. (2003). An EMIS for Palestine – the Education Management Information System in the West Bank and Gaza Strip. *Mediterranean Journal of Educational Studies*, 7(2), 61–92. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.520.1792&rep=rep1&type=pdf

The World Bank. (2017). *Sierra Leone—Multiple Indicator Cluster Survey 2017.* The World Bank. https://microdata.worldbank.org/index.php/catalog/3210

The World Bank. (2020). *School enrollment, primary, female (% gross)—Liberia | Data*. The World Bank. https://data.worldbank.org/indicator/SE.PRM.ENRR.FE?view=chart&locations=LR

World Health Organization. (2016). *Ebola data and statistics*. World Health Organization. https://apps.who.int/gho/data/view.ebola-sitrep.ebola-country-SLE-new-conf-prob-districs-20160511-data?lang=en

WorldPop. (2018). *Sierra Leone population 2020*. WorldPop. https://www.worldpop.org/geodata/summary?id=6429

van Wyk, C., & Crouch, L. (2020). *Efficiency and Effectiveness in Choosing and Using an EMIS*. UNESCO. http://tcg.uis.unesco.org/wp-content/uploads/sites/5/2020/09/EMIS-Buyers-Guide-EN-fin-WEB.pdf