Oxford Policy Management

CAMBRIDGE EDUCATION

esspin
Education Sector Support Programme in Nigeria

UKaid
from the British people

ESSPIN Composite Survey 3

Kwara State Report

Shefali Rai and Vanika Grover

7 October 2016

Acknowledgements

We thank the following people for their contributions to this report: Stuart Cameron has provided overall technical leadership and oversight of Composite Survey 3 (CS3), including inputs at various stages of producing this report. CS3 has been managed by Ifeatu Nnodu and Shefali Rai. Femi Adegoke of Oxford Policy Management (OPM) Abuja directed the survey team in Abuja. Ekundayo Arogundade and Kelechi Udoh, and their team of state coordinators and monitoring officers, managed the fieldwork. Babatunde Akano and Joshua Moriyonu led data management. David Megill provided sampling support and survey weights. Katharina Keck and Zara Majeed provided analysis of the data. Reg Allen analysed learner assessments, teacher assessments, and the teacher motivation survey. Alia Aghajanian provided analysis of the Annual School Census data. Rachita Daga and Sourovi De contributed analysis of School-Based Management Committee (SBMC) and inclusion indicators. Rachel Outhred and Tanya Lone worked on learner assessment analysis and redesign. Stuart Cameron provided helpful comments on an earlier draft of this report.

From ESSPIN, we acknowledge the contribution of Ron Tuck, who first devised the concept of the Composite Survey, as well as various leadership roles in CS1 and CS2 of Stuart Cameron, Kayode Sanni, John Kay, Lilian Breakell, Simon Thomson, Fatima Aboki, Sandra Graham and Jake Ross. For CS3, Abimbola Osuntusa, Fatima Aboki, Helen Pinnock, Kayode Sanni, Laura McInerney, Manjola Kola, Sandra Graham, Simon Thomson, Jake Ross, Ron Tuck, James Fadokun, Kingsley Ogbonna and Philip Adigun provided valuable information and support. Allan Findlay and Agbenu Onazi provided information on enrolment and school numbers from the annual school census.

The Composite Surveys would not have been conducted without UK aid funding provided by DFID, including the guidance and support of Barbara Payne, Emily Oldmeadow and Esohe Eigbike.

We are very grateful to the School Support Officers (SSOs), Social Mobilisation Officers (SMOs), and other local government and state staff who acted as data collectors for the Composite Survey. We are also very grateful to the state government officials who shared information on the context in which ESSPIN has been implemented. Finally, we are extremely grateful to all of the head teachers, teachers and students who kindly took the time to participate in this survey.

Executive summary

The Education Sector Support Programme in Nigeria (ESSPIN) (2008–17) seeks to improve learning outcomes for children of basic education age in six Nigerian states – Enugu, Jigawa, Kaduna, Kano, Kwara and Lagos. The aims of the ESSPIN Composite Surveys are to assess the effects of ESSPIN's integrated School Improvement Programme (SIP) and to report on the quality of education in the six ESSPIN-supported states. ESSPIN is funded by the UK Department for International Development (DFID) and managed by a consortium led by Cambridge Education. The Composite Survey has been carried out for ESSPIN by Oxford Policy Management (OPM).

This report presents findings for Kwara State from the first, second and third rounds of the Composite Survey (CS1, CS2 and CS3). These took place in 2012, 2014 and 2016 respectively. The surveys covered a wide range of indicators related to teachers, head teachers, School-Based Management Committees (SBMCs) and pupils. The aim of the surveys is to understand how schools in ESSPIN states are changing over time and whether schools that receive ESSPIN interventions are working better than those that do not. The main findings are as follows:

Head teacher effectiveness has not improved significantly since 2012. In 2016, 23% of head teachers in Kwara met ESSPIN's standard for effectiveness. Performance has improved on some fronts, such as the share of head teachers holding regular professional development meetings (which have been the focus of ESSPIN training). It has deteriorated on others, such as the share of head teachers taking action on teacher attendance. This is against a backdrop of teachers being unpaid for the five months leading up to and during the survey.

The share of schools that meet ESSPIN's criteria for **school development planning** effectiveness has improved from 6% in 2012 to 30% in 2016. Schools that have had greater exposure to ESSPIN's interventions in support of SBMCs are significantly more likely to meet the school development planning standard than those with less or no exposure to these interventions.

The share of schools that fully meet ESSPIN's **inclusiveness** standard fell from 78% in 2012 to 67% in 2016. Contrary to expectations, schools that have not received interventions in support of SBMCs are slightly more likely to meet the inclusiveness standard than those that have.

There have been large improvements in all indicators associated with **SBMC functionality** between 2012 and 2016. In 2016, 96% of SBMCs met ESSPIN's functionality standard, up from 30% in 2012. Schools that have received ESSPIN interventions in support of SBMCs were more likely to meet the standard than those that have not. The share of SBMCs that are deemed by ESSPIN to be inclusive of women and children improved between 2012 and 2016.

The share of **teachers** deemed competent by ESSPIN has fallen from 85% in 2012 to 76% in 2016. The high level of competence recorded in 2012 is likely to partly reflect the early years of the SIP in Kwara. Teachers' performance on the subject knowledge tests has worsened significantly since 2014. In line with this, only 30% of teachers met the strict version of ESSPIN's competence standard in 2016. ESSPIN-trained teachers performed better than non-ESSPIN-trained teachers on most indicators of teacher competence and on the literacy and numeracy tests.

The share of schools that meet ESSPIN's standard for **school quality** – measured as a combination of head teacher effectiveness, school development planning effectiveness, SBMC functionality and teacher competence – improved significantly in Kwara, from 11% in 2012 to 38% in 2016. As a result, an estimated 30,500 additional children in the state are now studying in good quality schools, as per this measure. Schools' performance on the strict version of this standard has stagnated. Schools that have received SBMC-related support performed better on almost all measures of school quality in 2016. They also recorded significantly faster improvements in their quality scores between 2012

and 2016. This suggests that ESSPIN's SBMC interventions have made a positive contribution to school quality in Kwara.

Pupils' learning outcomes in Kwara have worsened on three of the four assessments carried out as part of the Composite Surveys since 2012 and 2014. The exception is Grade 4 numeracy scores, which have improved slightly over this period. Learning outcomes are better in schools that have had received more support for SBMCs from ESSPIN. These differences are robust to controlling for school characteristics such as urban/rural location and distance from the Local Government Area (LGA). However, schools that have received such support have not seen faster improvements in their learning outcomes between 2012 and 2016.

Table 1: Kwara: Change over time – key indicators in 2012, 2014, 2016

	2012 (CS1)	2014 (CS2)	2016 (CS3)	Change 2012–16	Change 2014–16
Effective head teacher (%)	18.8	19.9	22.7	3.9	2.8
School development planning (%)	6.4	25	29.7	**+23.3***	4.7
Inclusive (%)	78.4	63.7	66.9	-11.5	3.2
Functioning SBMC (%)	29.6	75.1	89.8	**+60.2***	+14.7*
Competent teachers (%)	84.7	73.6	76.3	-8.4	2.7
Competent teachers (new measure, %)		38.4	30.1	n/a	-8.3
Good quality school (%)	11.4	26.8	37.7	**+26.3***	11
Good quality school (new measure, %)		17.8	16.2	n/a	-1.5
Grade 2 literacy score	495	510	479.9	-15.1	**-30.0***
Grade 4 literacy score	482.2	496.2	474.2	-8	**-21.9***
Grade 2 numeracy score	547.4	516.2	483.3	**-64.0***	**-32.8***
Grade 4 numeracy score	498.2	490.5	500.5	2.3	10
Note. * indicates statistical significance (p < .05)					

Table 2: Kwara: Key indicators in 2016, by years of ESSPIN Output 4 intervention

	None	Post-CS1	Pre-CS1	Estimated effect of one year of full intervention
School development planning (%)	4.7	17.9	45.1	**4.8***
Inclusive (%)	33.8	13.3	20.2	-4.4
Functioning SBMC (%)	68.6	90.2	99.8	6.1
Good quality school (%)	13.4	29.8	51.6	6.2
Good quality school (new measure, %)	8.7	5.5	22.9	0.8
Grade 2 literacy score	474.3	477.8	485.7	0.5
Grade 4 literacy score	466	486.5	478.1	3.5
Grade 2 numeracy score	470.8	489.7	493.1	3.3
Grade 4 numeracy score	470.9	522.8	520.3	**11.5***
Note. * indicates statistical significance (p < .05)				

Table 3: Kwara: Teacher competence, non-ESSPIN-trained versus ESSPIN-trained

	Non-ESSPIN-trained	ESSPIN-trained	Difference
Competent teachers (%)	62.3	85.4	+23.1*
Competent teachers (new measure, %)	24.4	33.7	+9.3
Teachers' English scale	484.2	497.3	+13.1
Teachers' mathematics scale	494.3	505.0	+10.7

Note. * indicates statistical significance (p < .05)

Table of contents

List of figures, tables and boxes

List of abbreviations

ACLED	Armed Conflict Location & Event Data Project
CAPI	Computer-assisted personal interviews
CBOs	Community-based organisations
CS1	Composite Survey 1
CS2	Composite Survey 2
CS3	Composite Survey 3
CSOs	Civil society organisations
DFID	UK Department for International Development
ESSPIN	Education Sector Support Programme in Nigeria
IRT	Item response theory
LGA	Local Government Area
LGEA	Local Government Education Authority
NGN	Nigerian Naira
OPM	Oxford Policy Management
PTR	Pupil-teacher ratio
SBMC	School-Based Management Committee
SDP	School Development Plan
SIP	School Improvement Programme
SMO	Social Mobilisation Officer
SSIT	State School Improvement Team
SSO	School Support Officer
SUBEB	State Universal Basic Education Board
UBEC	Universal Basic Education Commission

1 Introduction

ESSPIN, 2008–17, seeks to improve learning outcomes for children of basic education age in six Nigerian states – Enugu, Jigawa, Kaduna, Kano, Kwara, and Lagos. The ESSPIN Composite Surveys seek to assess the effects of ESSPIN's integrated SIP, and to report on the quality of education in the six ESSPIN-supported states. ESSPIN is funded by DFID, and is managed by a consortium led by Cambridge Education. The Composite Survey has been carried out for ESSPIN by OPM.

The first two rounds of the Composite Survey were carried out in 2012 and 2014. The surveys address five output indicators: teacher competence, head teacher effectiveness, school development planning, SBMC functionality, and inclusive practices in schools. They also address one outcome indicator, school quality, and one impact indicator, pupil learning achievement. The third round of the Composite Survey (CS3) collects comparable data on these indicators in order to provide information on the extent to which key school-level indicators in the six states have improved during the course of the programme.

This report focuses on the Composite Surveys' findings in Kwara State. It presents the key findings from CS3, compares these to the findings of the previous rounds of the survey, and draws out the implications of these findings for ESSPIN's contribution to school-level outputs and outcomes in the state.

1.1 ESSPIN's SIP

ESSPIN aims to bring about better learning outcomes for children of basic education school age in six states, with a range of activities at the state, national, local and school levels. It has four output streams that focus on

- strengthening federal government systems;
- increasing the capability of state and local governments as regards the governance and management of schools;
- strengthening the capability of primary schools to provide improved learning outcomes; and
- improving inclusion policies and practices in basic education (ESSPIN, 2013b).

Under the third of these outputs, ESSPIN's SIP aims to provide and support the use of structured materials that ensure teachers can deliver quality instruction, to strengthen teachers' understanding of literacy and numeracy concepts, and to improve academic leadership and school improvement planning by head teachers (Sanni, 2015).

The SIP typically works through a two-year modular programme of workshops and school visits conducted by local government School Support Officers (SSOs), after which schools continue to receive school visits from government officers to help maintain and continue to improve quality gains. At the same time, many of the same schools have received interventions under the fourth output stream, facilitating community involvement and inclusion through SBMCs.

1.2 ESSPIN in Kwara State

ESSPIN has worked with government primary schools in Kwara since 2009. A distinctive feature of ESSPIN's involvement in Kwara is that the SIP was rolled out to all public primary schools in the state right from the start. This is in contrast to the other five ESSPIN states, where the SIP was delivered to less than 10% of public primary schools in the first year of implementation. ESSPIN's

universal coverage of public primary schools in Kwara has meant that its contribution to school-level performance in the state cannot be inferred by comparing schools that have received ESSPIN support to those that have not.

ESSPIN's support to schools in Kwara has encompassed all three elements of the SIP: support to teachers, head teachers, and school improvement planning. An average of four teachers per school have received training on basic literacy teaching (early reading skills), basic numeracy teaching (number concepts, addition and subtraction), the use of teaching aids, classroom organisation, and the use of praise. Head teachers have received training on academic leadership, school planning, the management of teachers, and working with the community. This has been reinforced through regular monitoring and support visits by SSOs. In addition, some schools received two rounds of school grants to be spent on activities or investments included in their School Development Plan (SDP). The value of these grants has varied in line with school size: the average value was Nigerian Naira (NGN) 150,000 per year (ESSPIN, 2013b).

In addition to these interventions, since 2011/12 a subset of schools in Kwara have received support under ESSPIN's fourth output: improving inclusion policies and practices in basic education. ESSPIN has trained civil society members and government officers from the Department of Social Mobilisation – Social Mobilisation Officers (SMOs) – to enable them to train and mentor SBMCs. SBMC members, in turn, have been trained on: the roles and responsibilities of SBMCs; school planning and management; communication and leadership; change and relationships management; the participation of women and children in school improvement and education decision-making; resource mobilisation and financial processes; child protection and participation; and inclusive education and gender. This has been complemented by follow-up mentoring visits by SMOs.

The extent to which schools have received each of these interventions has varied from one year to the next (see Annexes B and C). Having received teacher and head teacher training for two consecutive years, schools in Kwara did not receive any such training in 2011/12 and 2012/13. This is in line with ESSPIN's approach in the state, which is based on the principle that training workshops represent a basic package of support to bring teachers and head teachers up to speed with how to do the key parts of their job. After this, the focus shifts to embedding that training through regular school visits carried out by SSOs – the focus here is on helping teachers apply their training to specific situations in their school or classroom. Some further training was, however, provided to head teacher, teachers, State School Improvement Teams (SSITs) and SSOs in 2013/14. This was geared towards addressing the gaps identified by CS1, and to bringing stakeholders up to speed on certain changes in the SIP – revisions to learning outcomes benchmarks, the distribution of lesson plans, and additional components related to inclusive education.

With regards to the intensity of training, the typical model in Kwara has involved six days of training for head teachers, three days of training for teachers, and 30 school support visits per year. A lower level of support was provided in 2015/16, with 15 school support visits made over the course of the year.

While all public schools in Kwara have received the same package of Output 3 interventions, the level of Output 4 support has varied across schools. As of 2016, 40% of schools in Kwara had received no Output 4 support, 18% had received some support since 2011/12, and the remaining 42% has received support since 2013/14 (see Annex C).

Since the SIP has been implemented at scale in Kwara right from the start, its delivery has always taken place through government structures and has not varied significantly over the course of the

programme. This is in contrast to the other five ESSPIN states, where delivery modalities have changed over time. ESSPIN's delivery model in Kwara has the following key elements:

- A core SSIT was created by ESSPIN in Kwara. This consists of state government staff (including lecturers and teachers) who sit within the State Universal Basic Education Board (SUBEB). The SSIT received training and support over a three-year period to enable it to develop the capacity of the Advisory Service Unit and SSOs, who in turn are in responsible for delivering the SIP. While the training delivered by SSITs is planned by ESSPIN, it is delivered – and also now paid for, managed and monitored – by the state.
- The SSOs trained by the SSIT deliver training to teachers and head teachers. They also carry out regular school visits to mentor teachers and support them to apply the content of training. ESSPIN programme staff support SSOs to deliver this type of mentoring and support.
- A similar model applies for Output 4 – ESSPIN staff develop the skills of civil society organisations (CSOs) and local government staff. CSOs are then contracted by the state to train SBMCs. They work in partnership with SMOs, who are local government employees, in civil society/government partnerships.

1.3 Contextual factors and their implications for the SIP in Kwara

This section describes key aspects of the backdrop against which ESSPIN's implementation has taken place in Kwara in recent years. These contextual factors are relevant when interpreting the changes in school-level outputs and outcomes between CS2 and CS3. While such changes may have resulted from ESSPIN support, they may also have been driven by other developments in the state over this period. This section considers the main developments in Kwara that may have positively or adversely influenced school-level outcomes in the state, or that may have interfered with the SIP's implementation in Kwara.

The main contextual change in recent years has been the sharp drop in oil prices, which has had major fiscal repercussions for almost all Nigerian states. The associated decline in federal revenue has affected funding for the Universal Basic Education Commission (UBEC), which in turn is the main source of revenue for basic education in the state and for the SIP. This has affected the monitoring of SIP activities – only 15 school visits were carried out in 2015/16, compared to 30 visits per year in previous years. Perhaps more importantly, the fiscal crunch has had significant indirect effects. At the time of CS3, teachers had not been paid for five months. This is likely to have had an adverse effect on teacher motivation, with potential implications for teacher attendance, time on task, and the extent to which teachers are making the effort to apply the new skills gained through ESSPIN training. It could also have had an adverse effect on head teachers' willingness to take action on teacher attendance (see Section 3.1).

The political context in Kwara has become slightly less favourable for ESSPIN in the aftermath of the 2015 elections. While the Governor was re-elected, the Commissioner of Education has changed and the SUBEB has been functioning without a new chair. Officially, education has remained a priority. However, there are indications that political support for basic education has fallen. This has been reflected in a decline in government-funded activities as part of the SIP in Kwara.

At the same time, there appear to have been some contextual changes that are supportive of school-level outcomes. The government is making some efforts to redress the shortage of teachers in rural areas by directly recruiting teachers from within understaffed rural communities. However, it is likely to take some time for the effects of this policy to work its way through the system.

Annual School Census data indicate that enrolment in public schools in Kwara was largely unchanged between 2009/10 and 2014/15 (Table 4). Annual School Census data also suggest that pupil–teacher ratios (PTRs) remained largely unchanged over this period, increasing very marginally from 15.7 in 2009 to 16.4 in 2014/15. Overall, it seems unlikely that changes in enrolment have had major implications for teaching quality between CS2 and CS3.

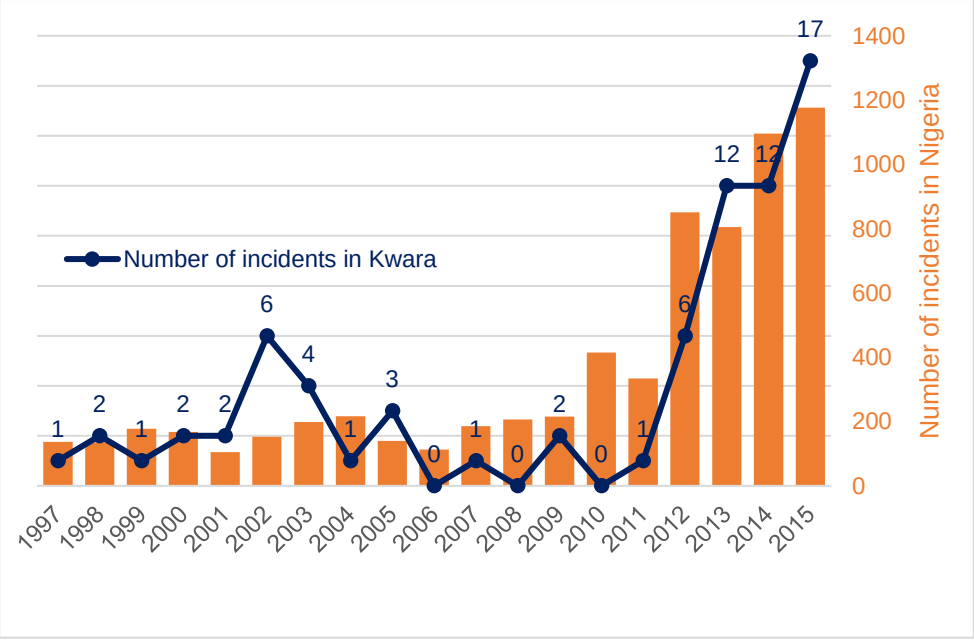
Table 4: Number of schools and enrolment in the 2009, 2013 and 2014 school censuses

	Enrolment	Number of schools	Enrolment change (%)
2009/10	199,604	1,448	
2013/14	198,759	1,497	-0.4
2014/15	199,868	1,528	0.6
Overall			0.1%

Note: Enrolment is for primary Grades 1–6.
Source: Annual School Census reports.

The incidence of political violence in Kwara has increased in recent years, but from a very low base (see Figure 1). Stakeholders reported that community clashes and attacks by armed herdsmen did lead to the temporary closure of a few schools in 2014, but things have returned to normal now. The implications for the school-level outcomes considered in the Composite Surveys are likely to be limited.

Figure 1: Incidents of political violence in Nigeria and Kwara



Source: Armed Conflict Location & Event Data Project (ACLED), Version 6 (1997–2015). Note all events from ACLED are included except for those categorised as protests which did not involve a fatality.

Table 5: Kwara – Political violence: Incidents and fatalities, 2010–2015

Variable	2010	2011	2012	2013	2014	2015
Events	-	1	6	12	12	17
Fatalities	-	0	5	26	11	11

2 Methodology and analysis

2.1 Evaluation strategy

2.1.1 Classifying the amount of ESSPIN Intervention

The original evaluation design for ESSPIN relied on maintaining a control group of schools which received no intervention, which could be compared to those with a longer history of intervention (those that began to receive ESSPIN support prior to 2012/13) and those where intervention started more recently (those that began to receive ESSPIN support in 2012/13 or after). However, as noted above, in Kwara the SIP has been implemented in all government primary schools since 2009. This uniformity means that it is not possible to make inferences about ESSPIN support under Output 3 by comparing outcomes in schools that have received different levels of ESSPIN support, as no such differences exist.

Certain comparisons can, however, be made. First, within a given school, only some teachers receive direct training from ESSPIN, while others do not. The latter group may, however, benefit indirectly from ESSPIN interventions, for instance through interactions with their peers or better pedagogical support from their head teachers. Outcomes can be compared across these two groups of teachers to make inferences about the effects of direct training by the SSOs. Second, schools have received different levels of support under Output 4. For indicators that are likely to have been influenced by this support, comparisons can be made between schools that have received varying levels of support from ESSPIN.[1] We classify schools in Kwara into three groups: those that have received no Output 4 intervention, those that began to receive some Output 4 interventions in 2011/12 or prior to this (referred to as pre-CS1 schools), and those that began to receive Output 4 support in 2012/13 or after (post-CS1 schools). Annex B contains details of the amount of Output 4 support provided to each group of schools over the last six years.

Table 6: Number of schools in Kwara by level of Output 4 intervention

Intervention group	Number of schools in Kwara
None	571
Post-CS1	644
Pre-CS1	263

Note: Schools are classified as pre-CS1 if they received at least five days of intervention of any kind under Output 4 in 2011/12 or prior to this.

2.1.2 T ypes of analysis

The purpose of CS3 is to provide insights into the changes over time in the states in which ESSPIN works, and to evaluate whether the ESSPIN model is having an effect in the schools in which its school improvement and community inclusion interventions have operated. We are interested in a wide range of output indicators: teacher competence, head teacher effectiveness, school development planning, school inclusiveness, and the functionality and inclusiveness of SBMCs. Some of these indicators are also combined to give an overall indicator of school quality. Finally, ESSPIN's impact is measured in terms of improved pupil learning outcomes, which we ascertain through test scores in numeracy and English literacy in Grades 2 and 4.

[1] A companion report, 'Composite Survey 3: Gender and Inclusion Report' (De and Cameron, 2016), focuses on ESSPIN's Output 4 interventions, which run in parallel with Output 3 and aim to improve inclusion and community participation in schools.

For each of these indicators, we analyse **changes over time** between CS1 and CS2, and between CS2 and CS3 for all schools in Kwara. We would expect schools in CS3 to have higher output, outcome and impact measures than schools in CS1 and CS2 as a result of improvements facilitated by ESSPIN's support. However, it is also difficult to disentangle the effects of ESSPIN's support from the impact of other changes that took place over this period. Differences between indicator performance in CS1, CS2 and CS3 cannot be attributed entirely to the ESSPIN intervention since there are other reasons why schools may be improving (or deteriorating) over time. In the chapters below, we refer to some of these other factors and their implications for interpreting the findings of the Composite Surveys.

Another caveat to note here is that CS1 does not represent an accurate baseline for ESSPIN's activities in Kwara since the SIP was first introduced in Kwara in 2009. This means that ESSPIN may have contributed to improvements in school-level outcomes prior to 2012, which would not have been recorded by the Composite Surveys. This adds to the challenges of drawing firm inferences about ESSPIN's impact by comparing changes in outputs and outcomes between 2012 and 2016.

In the sections below, we use statistical tests (t-tests) to give an indication of whether a difference in results over time is significant. Given that the Composite Surveys present findings for a sample of schools in the state, it is possible that differences in results are driven by the specific features of the sample of schools covered by the surveys, rather than broader trends across the entire population of schools. Significance tests provide an indication of whether a particular difference is likely to be driven purely by the specific features of the sample in question rather than the population of interest. If a particular difference is statistically significant, this means that we can say with a high degree of certainty that a corresponding difference does exist in the population of interest.

For indicators that are expected to have been influenced by ESSPIN's Output 4 interventions, we also look at **variations between the different intervention groups within the CS3 results**. We hypothesise that schools that have received more years of full ESSPIN intervention have higher output, outcome and impact measures than schools which have received fewer years of intervention.[2] To test this, we use a continuous measure of the years of full intervention that each school has received (one to five), and calculate the estimated effect of having received one additional year of intervention using a simple regression model with dummy variables for each state.

This approach allows us to come one step closer to estimating the effect of ESSPIN interventions, by controlling for one of the main confounding variables (the state). However, this will not be a conclusive indicator of ESSPIN's effect because there are also differences in school and pupil background characteristics within states (see Annex A). Controlling for this fully is a more difficult exercise, so we will only attempt this for our impact measure, pupil learning outcomes. In the case of Kwara, these measures can only provide an indication of the contribution of ESSPIN's Output 4 interventions. As noted above, all schools in the state have received a similar level of Output 3 support, so it is not possible to classify schools into different intervention groups when it comes to Output 3 interventions.

[2] As noted above, this type of comparison cannot be drawn for all indicators as all government primary schools in Kwara have received a similar level of Output 3 support.

2.2 Sampling, coverage and weights

2.2.1 Sample design

In CS3, all the schools visited in CS2 were visited again with the intention of collecting data that would enable us to draw inferences about what is happening in the population of schools across the six states, and within each state, through the use of sample weights. As in CS2, the sample design in CS3 prioritised the ability to draw conclusions across the six states, conceding that it would not always be possible to obtain statistically significant estimates within each state, given the high degree of variability in the types of schools that are found in some of the states, which makes it difficult to construct a representative sample. The sampling design also took account of the key aims of the study – to analyse change over time and differences between schools having received different amounts of ESSPIN intervention.

In Kwara, CS3 was carried out in 105 schools. The number of schools sampled in each of the categories (as defined in CS3, so taking account of the full period of intervention) is shown in Table 7.

Table 7: Sample in CS1, CS2 and CS3 and population of schools, by state and with intervention groups

	Output 4 intervention category	CS1 sample (2012)	CS2 sample (2014)	CS3 sample (2016)	Population
Kwara	None	14	14	14	577
	Post-CS1	34	34	34	644
	Pre-CS1	53	56	56	263
	Total	102	105	105	1485

Note: The sample size shown is the actual sample for which data were collected. Intervention groups reflect the number of years of intervention the schools had received by the end of the 2014/2015 school year. For one school in the sample, the Output 4 intervention group is not known. As a result, the sum of schools in the three intervention groups does not add up to the total number of schools in the sample.

Within each school, the survey team conducted interviews with the head teacher, the SBMC chairperson or deputy, teachers and pupils. As in CS2, six teachers were sampled. However, for CS3 we attempted to find the six teachers interviewed during CS2, using their photographs and name information and to interview them again so as to be able to assess changes over time as well as rates of change in teacher competence and test results with more precision. In Kwara, we were able to track 55.5% of the CS2 teacher sample. The main reasons why teachers could not be tracked were that they no longer taught at the same school or were not present on the day.[3]

In schools where fewer than six teachers from CS2 were re-sampled, the sample was topped up to six teachers by randomly selecting teachers from the head teacher's register. Fieldwork teams asked head teachers to complete such a register in cases where this had not already been done for the day of the visit. Team supervisors entered the number of eligible teachers into the computer-assisted personal interviews (CAPI) system, which then randomly selected the additional top-up teachers to be sampled. We excluded teachers who only teach religious subjects and those that do not teach in Grade 1 – 6.

[3] In addition, in a few cases teachers no longer taught in Grades 1–6 or no longer taught non-religious subjects, and were therefore ineligible. Teachers could also not be tracked if the school was replaced during CS3.

As in CS2, 16 pupils were sampled per school in CS3, with a sample of four pupils per school for each of the tests (Grade 2 literacy, Grade 2 numeracy, Grade 4 literacy and Grade 4 numeracy). For CS3, we attempted to sample pupils who are currently being taught by one of the sampled teachers to allow us to better link teachers with pupils in the analysis, and examine, for example, whether pupils with better learning outcomes are taught by more competent teachers. To do this, we first sampled teachers in CS3. While sampling them, we gathered information from the head teacher on which teachers teach which arms of Grade 2 and 4, filling in an 'arm eligibility grid'. From this, field teams determined the 'eligible arms' (i.e. arms where at least one of the sampled teachers was teaching) and sampled learners only from these arms. To ensure that the full sample size was maintained, the sample was topped up if there were fewer than eight learners in the eligible arms. In addition, if none of the sampled teachers taught in Grade 2 (or Grade 4), learners were randomly selected from all arms in that grade.

In CS2, learners were sampled from the learner attendance register (if available), but these were found to not always be accurate. For CS3, each learner in the eligible arms was given a 'sampling card', which was either blank or contained a number. Learners who were handed cards with even numbers were sampled for the numeracy test; learners with odd numbers were sampled for the literacy test. This ensured that the sampling was done from the learners actually present in the class rather than from the potentially incomplete pupil register.

Although it would have been useful to trace the same pupils over time, this was not seen as feasible because, for the children sampled in CS1, we only have their names, which is not always sufficient to identify the same children three to four years later. We therefore collected a random sample within each school in CS3 as well as in previous rounds of the survey.

In addition to the main sample of 16 pupils and six teachers, an additional four pupils and two teachers were selected in each school by the CAPI system, as 'replacements'. Replacements were included in the survey in cases where teachers and pupils from the main sample turned out not to be available at the school, despite having been recorded as present in the register. Replacements could not be used in any other circumstances, however. In practice the option of replacing teachers and pupils was used rarely.

A number of schools were found to operate double shifts, with some classes taught in the morning and others in the afternoon. Double shift schools where different teachers taught in both shifts were sampled separately. We sampled morning teachers from teachers who were present, and afternoon teachers from the teacher attendance record for the previous day's afternoon. If the sampled teachers did not turn up in the afternoon, we used replacements. We sampled pupils from the arms taught by these sampled teachers. If this included arms taught in the afternoon, we sampled from the children who attended the previous afternoon, according to the pupil register.

In double shift schools where the same teachers taught both morning and afternoon on the day of the visit we sampled teachers who were present in the morning. We sampled pupils from the arms taught by these sampled teachers. If these included arms taught in the afternoon, we sampled from the children who attended the previous afternoon, according to the pupil register.

Table 8: CS3 – Intended survey sample in Kwara

Targeted sample size						
State	Schools	Head teachers	SBMC members	Teachers	P2 pupils	P4 pupils
Kwara	105	105	105	630	840	840

2.2.2 Weights

Simple averages of the results from the Composite Survey data would not be representative of what is happening across the state, because the profile of schools included in the survey is not identical to the profile of schools in the state as a whole. We address this by applying sample weights which give greater weight to the results obtained from schools that are relatively under-represented in the survey. Sample weights were calculated for the CS1, CS2 and CS3 schools, teachers, and pupils. A smoothing technique was also applied to reduce the variability of the weights and to avoid the design effects problem encountered in the CS1 analysis (see Megill, 2014b).

Most of the analysis that follows applies weights to sample statistics calculated within each round and intervention group, which can then be used as estimates of the whole population of schools in the state. However, part of the analysis compares change within individual school over time. For this we are limited to the set of schools which were sampled at each of the time points over which the comparison is conducted (e.g. an analysis of change in individual schools between CS2 and CS3 is limited to those schools included in both the CS2 and CS3 rounds of the survey). Additional sets of weights were calculated for use with these 'panels' of schools.

In addition, because we re-sampled teachers from CS2, in some of our analyses, we compare how the same teacher performed in CS3 compared to CS2. For this we are limited to the set of teachers who were sampled during both CS2 and CS3. Another set of weights was calculated for use with this panel of teachers.

2.2.3 Sample coverage

Table 9 lists the number of respondents covered by each of the survey instruments, and how this compares to the number of targeted respondents in each category. In line with the targeted sample, the field team visited 105 schools in Kwara. Head teachers and SBMC chairs/deputies were interviewed in all the schools visited. The number of teachers interviewed fell short of the targeted sample size. The primary reason for this was that many schools had fewer than six eligible teachers (i.e. teachers who teach non-religious subjects to Grades 1–6). Sample coverage of *eligible* teachers in Kwara was 99.6%. There were some further minor differences between the numbers of teachers interviewed and those who were tested or whose lessons were observed because some teachers did not give their consent to be tested or observed; and because some teachers reportedly became ill and so could not complete the test or lesson. Some teachers were also not able to stay for the teacher tests, which were held after school hours. Similarly, the number of pupils assessed fell slightly short of the targeted number because some schools had fewer than eight pupils in P2 or P4.

Table 9: Sample coverage in CS3

	Targeted sample size	Number of respondents covered	Sample coverage (% of targeted sample size)
Number of schools	105	105	100%
Head teacher interview	105	105	100%
SBMC interview	105	105	100%
Teacher interview	630	495	79%
Teacher literacy test	630	498	79%
Teacher numeracy test	630	498	79%
Lesson observations	630	492	78%
P2 literacy	420	385	92%
P2 numeracy	420	373	89%
P4 literacy	420	380	90%
P4 numeracy	420	366	87%

2.3 Training, pilots and fieldwork model

Fieldwork for CS3 was conducted using CAPI during April to June 2016. We made a number of changes to instruments to take on board some additional concerns and to make use of innovations introduced in other recent Nigerian school surveys. At the same time we retained the questionnaire items required for comparability with previous rounds of the Composite Survey. Changes made to the instruments included the following:

- The addition of a scale for measuring teacher motivation and the quality of interaction between teachers, which has been used in other recent Nigerian schools surveys.
- Additional items designed to assess school management more broadly, including items on actions taken to increase enrolment, issues related to teacher attendance, actions to make sure teachers are present in classrooms during lessons, and actions to improve school materials or infrastructure. We also visually inspected whether teachers and learners were in class around the beginning of the day and after the long morning break.
- The overall technical report on CS2 (Cameron, 2015a) noted that there were apparently large increases in enrolment in some states, along with increases in pupil–classroom ratios and pupil–teacher ratios. To better understand enrolment changes we added questions in the head teacher questionnaire to clarify the number of children currently enrolled in the school and the number of teachers. Head teachers were also asked about changes in enrolment that had occurred compared to the previous year, and where they reported increases in enrolment, what issues it might have caused and how they had dealt with the change.
- An expanded version of the wealth index included in the Grade 4 pupil tests, to better control for children's home backgrounds.

As in CS2, pupil assessments in CS3 were administered using CAPI. Children were given a printed pupil book to read and write in. The interviewers made use of a tablet computer, which prompted them on the questions the children were to be asked orally, gave instructions on the administration of the different test items, including timing, and allowed them to input whether each part of each question was answered correctly or incorrectly (or not attempted at all) by the pupil. A number of changes were made to the CAPI systems and manuals for the administration of the pupil tests, to make them easier to train on and administer. This included a clear manual with consistent

instructions across questions of a particular type, automated timers for timed questions, and translations into Hausa, Igbo and Yoruba of text that did not need to be read in English.

The instruments were pre-tested over two days in Abuja during April 2016. State coordinators and monitoring officers collected the data on CAPI after they had been trained on the instruments. Minor revisions were made to the instruments in consultation with state coordinators.

Table 10 lists the instruments used in CS3, together with the indicators relevant to outcomes, outputs or impact that were gathered from each instrument. The instruments were also used to gather intervention information, such as whether individual teachers had received ESSPIN training or not, and pupil-level information on socio-economic status, age, language spoken at home, and gender.

The process of revising instruments for CS3 does leave some possibility of measurement error in comparisons between the previous Composite Surveys and CS3. We tried to ensure consistent and manageable data collection within CS3, by setting clear guidance for data collectors through detailed data collection manuals, constant oversight, and intensive training for all data collectors, including three pilot field days. Although we avoided large changes in instruments that would compromise comparability with CS1 and CS2, any change in questionnaire format or wording, training, and data collection procedures can potentially affect the results. However, since changes in measurement are consistent across the different intervention categories, they should not affect any within-CS3 comparisons.

Table 10: Instruments used in CS3

Instrument	Outcome / output / impact indicators
Structured interview with head teacher	Number of lesson observations during past two weeks; number of professional development meetings this school year; teacher attendance book; actions by head teacher to promote teacher attendance and improve pupil attendance; written evidence of school self-evaluation process for school year; SDP for school year available; activities relating to strengthening teaching and learning in the SDP; activities relating to improving access in the SDP; evidence of activities in the SDP being carried out; up-to-date cashbook.
Structured interview with SBMC chairperson and members	Number of SBMC meetings this school year; SBMC awareness-raising activities; steps taken by SBMC to address exclusion; SBMC networking with community-based organisations (CBOs), traditional or religious institutions, other SBMCs, and Local Government Education Authorities (LGEAs); SBMC has a women's committee and a children's committee, and how often these committees meet; SBMC has contributed resources to the school; visits by the SBMC to the school this school year; number of SBMC meetings attended by at least one woman and by at least one child; issues raised by female and child members; action taken on issues raised by female and child members; whether children's committee had a trained facilitator; action for commonly excluded groups; SBMC raised issue of children's exclusion.
Structured interview with teacher	Knowledge of English and maths curriculum benchmarks; school opening time.
Lesson observation	Number of forms of classroom organisation used; number of teaching aids used; number of times teacher praised or reprimanded children; participation of children from different zones of the classroom; participation of boys and girls in the lesson.
Teacher tests	Teacher test scores in English literacy and numeracy.
Pupil tests	Pupil test scores in English literacy and numeracy at Grades 2 and 4.
General observation	Length of morning break; number of classes where pupils and teachers are in class within half an hour of starting time and long morning break.

3 School management and head teachers

ESSPIN's interventions include leadership training for head teachers on managing the school and its teachers, planning for the school's development, advocating for more resources, and ensuring that the school is inclusive. ESSPIN also supports the development of SBMCs. This includes training and mentoring on how SBMCs can encourage the participation of women and children. This chapter examines how well schools in Kwara are doing on each of these fronts.

ESSPIN's logframe identifies and defines a number of indicators related to school management, inclusiveness and SBMCs. The logframe groups these indicators into a set of 'standards' or composite indicators. These are as follows:

- Head teacher effectiveness: A head teacher is deemed to be effective if they engage in a set of practices including observing teachers' lessons, holding professional development meetings with teachers, monitoring teacher attendance, keeping records, and ensuring that the school adheres to a regular schedule.
- School development planning: As part of the SIP, schools are encouraged to carry out a self-review process involving the head teacher, teachers, SBMCs, parents and other community members. The aim of this process is to identify the school's strengths and weaknesses, and then list the steps that need to be taken to improve it in an SDP. The SDP can also be used to request resources from local government or the community. The associated logframe standard assesses whether a self-evaluation has been carried out, whether the school has an SDP, and whether it has implemented the activities in its SDP.
- School inclusiveness: This refers to the extent to which the school makes an effort to include all learners, regardless of gender or socio-economic background. Inclusiveness is assessed on the basis of the steps listed in the SDP and actions taken to boost access, as well as the extent to which teachers encourage the participation of all children in the classroom.
- SBMCs' functionality and performance: The associated standards assess the extent to which SBMCs are functioning and active, and the degree to which they ensure that women and children are actively participating in their activities.

The rest of this section describes each of these standards and then presents associated findings from the Composite Surveys.

3.1 Head teacher effectiveness

Box 1: Head teacher effectiveness: Key findings

- In 2016, 23% of head teachers in public primary schools in Kwara met ESSPIN's standard for effectiveness. This indicator has not improved significantly since 2012 or 2014.
- Performance has improved on some fronts, such as the share of head teachers holding regular professional development meetings, but deteriorated on others, such as the share of head teachers taking action on teacher attendance.

ESSPIN defines head teacher effectiveness with regards to seven criteria set out in its logframe (see Box 2). The first two criteria relate to the pedagogical support that head teachers provide to teachers, the next relates to the steps that head teachers take to boost teacher attendance, and the final four relate to school management practices that have implications for time on task.

Box 2: Logframe criteria for head teacher effectiveness

A head teacher must ensure that five out of seven of the following criteria are met in order to meet the head teacher effectiveness standard:

1) carried out two or more lesson observations in the past two weeks;
2) held four or more professional development meetings since the start of the 2014/15 or 2015/16 school year (NB: the survey took place more than nine months into the school year);
3) school has a teacher attendance book and the head teacher recalls at least two actions taken to promote teacher attendance;
4) clear school opening time: more than 50% of pupils sampled agree on the school opening time and more than 50% of teachers sampled agree on the school opening time;
5) more than 50% of classes are in their classroom with their teacher within 30 minutes of school opening time;
6) length of morning break is 35 minutes or less; and
7) more than 50% of lessons observed finished within five minutes of a standard 35-minute lesson duration (i.e. the lesson was between 30 and 40 minutes long).

3.1.1 Changes in state-level outcomes between CS1, CS2 and CS3

There has been no significant change in the share of head teachers that met ESSPIN's standard for effectiveness between 2012 and 2016 (see Table 11). Over this period, some indicators have improved, notably the share of head teachers holding regular professional development meetings with teachers and the share of schools in which at least half of lessons are of an appropriate length. However, this has been offset by a sharp decline in the share of head teachers taking action on teacher attendance. This is particularly notable given that the teacher absenteeism rate has risen more than two-fold since 2014 to 25% in 2016 (see Table 11). This may be related to the five month delay in salary payments noted above, which may have undermined head teachers' motivation and willingness to take action on teacher attendance. The share of schools with a clear opening time has also fallen sharply since 2012, although it is unclear how meaningful this is as an indicator of school management. Field observations suggested that children were confused about whether to consider the time that they arrived at the school, the time of assembly, or the time when lessons started, as the school opening time.

Overall, in 2016 the share of head teachers meeting ESSPIN's effectiveness standard remained fairly low, at 23%. While some indicators have improved over this period, it is notable that seven years after ESSPIN was first introduced in Kwara, the prevalence of certain practices that the SIP has promoted remains limited. For instance, only 34% of head teachers carry out regular lesson observation. The share of head teachers who hold regular professional development meetings has doubled since CS1, but at 38% remains lower than expected.

Table 11: Kwara: Head teacher effectiveness in CS1, CS2 and CS3

	2012 (CS1)	2014 (CS2)	2016 (CS3)	Change: 2012 vs. 2016	Change: 2014 vs. 2016
(1) Lesson observations (%)	24.6	17.9	34.1	+9.5	+16.2
(2) Professional development meetings (%)	20	37.6	38.2	+18.2*	+0.6
(3) Action on teacher attendance (%)	73.3	45.7	39.7	-33.7*	-6.0
(4) Clear opening time (%)	43	16	9.6	-33.4*	-6.4
(5) In class on time (%)	89.2	74.1	85.4	-3.8	+11.3
(6) Appropriate morning break (%)	88	78.8	78.9	-9.1	+0.1
(7) Appropriate lesson length (%)	15	27.3	45.1	+30.0*	+17.8
Number of criteria fulfilled (/7)	3.5	3	3.2	-0.3	+0.2
Effective head teacher (5/7 criteria met) (%)	18.8	19.9	22.7	+3.9	+2.8
Additional indicators					
In class on time after break (%)	n/a	n/a	87.4	n/a	n/a
-- Classes where learners are present on time (%)	n/a	n/a	99.3	n/a	n/a
-- Classes where teachers are present on time (%)	n/a	n/a	83.6	n/a	n/a
-- Classes where learners and teachers present on time (%)	n/a	n/a	83.6	n/a	n/a
Teacher absenteeism (%)	n/a	10.9	25.1	n/a	+14.2*
* Indicates that change over time is statistically significant (p < 0.5)					

These findings raise a number of questions for ESSPIN. In particular, it is worth exploring why head teachers have become less likely to take action on teacher attendance. One possibility is that they have limited efficacy to influence teacher attendance (partly because they lack the authority to take disciplinary action against teachers). Another is that their motivation and sense of authority has been undermined by the fiscal crisis and associated delays in salary payments. Similarly, it is worth exploring why progress has stalled between 2014 and 2016 in regard to the share of head teachers carrying out regular professional development meetings.

There is one further caveat that is worth noting when interpreting these results. Criterion 7 in the table above (appropriate lesson length) records whether at least 50% of lessons observed finished within five minutes of a standard 35-minute lesson duration. However, ESSPIN has been trying to persuade states to move to one-hour lessons for English and mathematics, to allow them to more effectively deliver the lesson plans that it has designed and distributed to school. This criterion is, therefore, at odds with recent developments as part of ESSPIN. If we consider a slightly different indicator that only penalises schools for having lessons that are too short, we see quite different findings to those recorded in the table above. In particular, the share of schools in which at least 50% of lessons were at least 30 minutes long stood at 90% in 2014 and 93% in 2016. One complication in regard to interpreting these findings is that ESSPIN has only promoted one-hour lessons for English and mathematics, whereas the lesson observations have encompassed a broader range of subjects.

3.2 School development planning

Box 3: School development planning: Key findings

- Between 2012 and 2016 there was a large and significant improvement in the share of schools that met ESSPIN's logframe criteria for school development planning effectiveness. Despite this improvement, only 30% of schools met the standard in 2016.
- Schools that have had greater exposure to ESSPIN's Output 4 interventions are significantly more likely to meet the school development planning effectiveness standard than those with less or no exposure to these interventions.

ESSPIN's leadership training encourages and supports head teachers to review their school's performance and put together an SDP, which can then be used to advocate for resources from the local government or the community. ESSPIN encourages schools to include a range of measures in their SDPs that go beyond investments in the school's infrastructure and include other measures to strengthen teaching and learning and promote access. Head teachers are also trained on using a cashbook to record the school's expenditure and income. It is expected that these measures will support the effectiveness of school development planning. ESSPIN assesses this on the basis of five criteria, outlined in Box 4.

Box 4: Logframe criteria for the effectiveness of school development planning

The school must meet criterion 1 and criterion 2 listed below, and at least two out of three of the remaining criteria, in order to meet the effective school development planning standard:

1) written evidence of school self-evaluation process for current school year;
2) SDP for current school year available;
3) SDP contains three or more activities which aim to strengthen teaching and learning;
4) physical evidence of four or more activities from SDP having been carried out; and
5) cashbook is up-to-date (balanced in the last 60 days).

Between 2012 and 2016, there was a large and significant improvement in the share of schools that met ESSPIN's logframe criteria for school development planning effectiveness. In particular there were large improvements in the share of schools that have carried out a self-evaluation, prepared an SDP, and have three or more measures to strengthen teaching and learning in their SDP. Between 2014 and 2016 the only criterion for which significant change was recorded was the share of schools that were able to provide written evidence of a self-evaluation, which continued to increase.

Table 12: Kwara: SDP effectiveness in CS1, CS2 and CS3

	2012 (CS1)	2014 (CS2)	2016 (CS3)	Change: 2012 vs. 2016	Change: 2014 vs. 2016
(1) Written evidence of school self-evaluation process (%)	29.8	47.2	67	+37.2*	+19.8*
(2) SDP available (%)	43.5	80	82.3	+38.7*	+2.3
(3) SDP contains three or more activities to strengthen teaching and learning (%)	24.6	46.3	44.7	+20.0*	-1.7
(4) Evidence that four or more activities from SDP carried out (%)	11.2	21.3	20.9	+9.7	-0.5
(5) Cashbook up-to-date (%)	13.8	25.4	26.1	+12.3	+0.6
Number of SDP criteria fulfilled (/5)	1.1	2.2	2.4	+1.3*	+0.2
School meets effective school development standard (%)	6.4	25	29.7	+23.3*	+4.7

* Indicates that change over time is statistically significant ($p < 0.5$)

Despite the nearly five-fold improvement seen between 2012 and 2016, just under one-third (30%) of schools met ESSPIN's effective school development standard in 2016. This is driven by weak performance on two criteria: the presence of an up-to-date cashbook, and evidence that four or more activities from the SDP have been carried out. This suggests that head teachers may need further support and training on financial management. It also highlights the need to engage with the reasons why there is limited implementation of activities in the SDP. For instance, it is possible that some schools prefer to focus on a smaller number of high-impact measures (raising questions about whether the focus on four or more activities being implemented is justified). Difficulties with mobilising resources from the LGA may also be partly responsible for limited implementation of SDPs.

We also analyse SDP effectiveness by intervention group in 2016. While the process of school development planning is led by the head teacher, the SBMC is also supposed to be involved in this process. As a result, there could be some link between the amount of SBMC-related support that schools receive and the efficacy of the planning process.

Table 13 highlights that schools that had received ESSPIN's Output 4 interventions were significantly more likely to meet the SDP effectiveness standard in 2016 than those that did not receive such support. Among schools that have received Output 4 support for the longest duration (pre-CS1 schools), 45% met the effectiveness standard in 2016. In contrast, only 5% of schools that received no Output 4 support and 18% of schools that have received such support after CS1 met the standard. These findings indicate that the likelihood of meeting this standard increases by 4.8 percentage points for each year of Output 4 intervention. As noted above, this is not an estimate of ESSPIN's causal impact but lends support to the hypothesis that longer exposure to ESSPIN's Output 4 intervention has a positive effect on school development planning.

Table 13: Kwara: SDP effectiveness in 2016, by intervention group

	none	post-CS1	pre-CS1	Estimated effect of one year of full intervention
(1) Written evidence of school self-evaluation process (%)	39.7	59.3	82.4	5.7
(2) SDP available (%)	82.7	89.1	80.1	-0.1
(3) SDP contains four or more activities to strengthen teaching and learning (%)	27.5	53.9	50.5	4.3
-- No. activities in SDP to strengthen teaching and learning	2.3	2.8	2.2	0.0
(4) Evidence that four or more activities stated in SDP carried out (%)	11.1	9.8	28.5	1.3
-- No. activities in SDP carried out	1.6	1.3	2.3	-0.0
(5) Cashbook up-to-date (%)	7.2	21.9	36.2	4.3*
-- School has a cashbook (%)	70.5	62.1	74.2	-1.8
Number of SDP criteria fulfilled (/5)	1.7	2.3	2.8	0.2
School meets effective school development planning standard (%)	4.7	17.9	45.1	4.8*

* Indicates that change over time is statistically significant ($p < 0.5$)

3.3 School inclusiveness and SBMCs

Box 5: School inclusiveness: Key findings

- The share of schools that fully met ESSPIN's inclusiveness standard fell from 78% in 2012 to 67% in 2016.
- Contrary to expectations, schools that have not received ESSPIN's Output 4 interventions are slightly more likely to meet the inclusiveness standard than those that have.
- In 2016, 96% of SBMCs met ESSPIN's functionality standard, up from only 30% in 2012. There were improvements in all aspects of SBMCs' performance over this period.
- Schools that have received Output 4 support were more likely to have functional SBMCs.
- The share of schools meeting the women's and children's inclusiveness standards increased between 2012 and 2016, although the latter remains low, at 26%.

3.3.1 School inclusiveness: Meeting the needs of all pupils

The criteria on school inclusiveness measure the extent to which the school makes efforts to include all learners, including those from disadvantaged backgrounds. ESSPIN's overall standard for school inclusiveness is based on four criteria (Box 6). Further detail on these is provided in the companion Gender and Inclusion Report.

Box 6: Standard for school inclusiveness

The school must meet at least three of the four criteria listed below in order to meet the school inclusiveness standard. The standard is partially met if two criteria are met:

1) head teacher states three or more actions that he/she has taken to improve pupil attendance;
2) SDP contains two or more activities which aim to improve access;
3) more than 50% of teachers observed provided evidence of using two or more assessment methods (marked class test, marked pupil workbook, or graded examination paper); and
4) more than 50% of teachers observed met the spatial inclusion criterion (defined as engaging with at least one pupil from four different areas of the classroom during a lesson) and more than 50% of teachers observed met the gender inclusion criterion (defined as engaging with boys and girls proportionally to their presence in the classroom within a 10% margin; for example, if the class contains 50% girls then teachers who engage with girls in between 60% and 40% of total engagements meet the criterion).

The share of schools that fully met ESSPIN's inclusiveness standard fell between 2012 and 2014 and has only recovered slightly since then. Between 2012 and 2016 some inclusiveness indicators improved while others worsened. In particular, the share of schools that had more than two actions in their SDP to improve access for disadvantaged children improved sharply. This is consistent with broader trends in the effectiveness of school development planning. Meanwhile, the share of head teachers taking three or more actions to improve pupil attendance fell sharply. The reasons for this are unclear and would be worth exploring further.

Table 14: Kwara: School inclusiveness in CS1, CS2 and CS3

	2012 (CS1)	2014 (CS2)	2016 (CS3)	Change: 2012 vs. 2016	Change: 2014 vs. 2016
(1) Three or more actions to improve attendance	76.4	24.2	25.1	-51.2*	+1.0
Number of actions on learner attendance	3.4	2.1	2.1	-1.4***	-0.1
(2) Two or more activities in SDP to improve access for disadvantaged children	14.7	47.8	55.5	+40.8*	+7.8
Number of activities on access	0.4	1.7	1.5	+1.1***	-0.1
(3) >50% of teachers use two or more assessment methods	78.3	91.8	72.7	-5.5	-19.1
(4) >50% of teachers spatially inclusive and >50% are gender inclusive	52.1	17.9	33.4	-18.8*	+15.5
Number of inclusiveness criteria fulfilled (/5)	2.2	1.8	1.9	-0.3	+0.0
Inclusiveness score	61	55.7	39.1	-21.9*	-16.6*
School partially met inclusiveness standard	44.7	17.8	22.8	-21.8*	+5.0
School fully met inclusiveness standard	78.4	63.7	66.9	-11.5	+3.2

* indicates that change over time is statistically significant ($p < .05$)

Table 14 also includes an alternative measure of inclusiveness, which points to a similar deterioration over time in the performance of Kwara's schools. This 'inclusiveness score' is a percentage score based on the number of actions taken to improve attendance, the number of activities in the SDP on access, the average number of assessment methods used, the average

number of zones participating in each lesson observed (observers imagined the classroom as being divided into six zones), and a measure of the extent to which girls and boys participated equally in the class. It seeks to assess changes in the underlying behaviour of schools and is less sensitive to the specific cut-offs used in the ESSPIN logframe criteria.

Contrary to expectations, schools that have not received ESSPIN's Output 4 support were slightly more likely to fully meet the inclusiveness standard in 2016 than those that did receive such support. The inclusiveness score (described above) was almost the same across the three groups of schools. However, there is one caveat to note when interpreting this – many elements of the inclusiveness standard are linked to teacher and head teacher training, which is provided as part of Output 3 and has been received by all schools in Kwara. As a result, the estimated intervention effects in the table below are a weak indicator of ESSPIN's contribution to school inclusiveness in Kwara.

Table 15: School inclusiveness in 2016, by intervention group

	None	Post-CS1	Pre-CS1	Estimated effect of one year of full intervention
(1) Three or more actions on learner attendance (%)	33.3	10.4	25.3	-3.1
-- Number of actions on learner attendance	2.2	1.9	2	-0.1
(2) Two or more activities in SDP on access (%)	44.9	74.6	55.3	4.3
-- Number of activities on access	1.2	2.1	1.6	0.2
(3) >50% of teachers use two or more assessment methods (%)	79.2	56.9	74	-4.2
(4) >50% of teachers spatially inclusive and >50% are gender inclusive (%)	53.1	19.2	27.9	**-7.5***
Number of inclusiveness criteria fulfilled (/4)	2.1	1.6	1.8	-0.1
Weighted sum inclusiveness score	38.7	37.4	39.7	-0.2
School fully met standard (3–4 criteria) (%)	33.8	13.3	20.2	-4.4
School partially met standard (2–4 criteria) (%)	68	62.9	67.4	-1.4
Additional indicators				
Enrolment increased since last year (%)	32.9	28	73.4	1.9
Change in enrolment since last year	-0.1	0.2	0.1	0.1

* indicates that estimated effect of one year of full intervention is statistically significant (p < .05)

3.3.2 SBMCs

SBMCs are considered by ESSPIN to be functioning well if they meet regularly and work with the community, traditional and religious institutions and local government to address the school's needs, raise resources for the school, and find ways to tackle exclusion. They are expected to have a women's committee and a children's committee, and to keep financial records. They are also expected to play a supervisory role, marked by regular visits to the school by the chairperson and other SBMC members. In line with this, ESSPIN uses nine criteria to assess SBMC functionality (see Box 7). SBMCs are considered to be effective if they meet at least five of the nine criteria. In most cases, these require evidence to be presented, rather than just accepting the word of the respondent (usually the SBMC chairperson). As a result, they partly reflect the ability of the SBMC to keep good records of their activities.

Box 7: Logframe criteria for SBMC functionality

The school must meet at least five of the nine criteria listed below in order to meet the SBMC functionality standard for the current school year:[4]

1) two or more SBMC meetings have taken place since the start of the current school year (written evidence);
2) SBMC conducted awareness-raising activities (written or oral evidence);
3) SBMC took steps to address exclusion (written or oral evidence);
4) SBMC networked with CBOs, traditional or religious institutions, or other SBMCs (written or physical evidence);
5) SBMC interacted with LGEAs on education service delivery issues (written or physical evidence);
6) an SBMC women's committee exists (written or physical evidence);
7) an SBMC children's committee exists (written or physical evidence);
8) SBMC contributed resources for the school (written or physical evidence); and
9) SBMC chair visited the school at least three times since the start of the current school year (written evidence).

The share of SBMCs that meet ESSPIN's functionality standard has increased from 30% in 2012 to 96% in 2016. There have been large improvements in all of the indicators associated with SBMC functionality over this period. This trend began in 2012–14 and has continued since then. The vast majority of SBMCs in Kwara now conduct awareness-raising, network with other organisations, hold regular meetings, contribute resources to schools, and take steps to address exclusion. The main fronts on which scope for improvements remains are the extent to which SBMCs make regular visits to schools and interact with LGEAs (criteria 5 and 9 in the table below). The relatively low share of SBMCs that interact with LGEAs may reflect broader systemic issues – such as the extent to which LGEAs are willing to engage with SBMCs and their perceived responsiveness to the latter's needs – which would call for measures that go beyond further training and mentoring of SBMCs.

[4] A slightly different standard, with 10 criteria, was used in CS1. The new standard, with nine criteria, was applied to both the CS1 and CS2 data.

Table 16: Kwara: SBMC functionality in CS1, CS2 and CS3

	2012 (CS1)	2014 (CS2)	2016 (CS3)	Change: 2012 vs. 2016	Change: 2014 vs. 2016
(1) Two or more meetings this school year	41.3	71.7	83	+41.6*	+11.3
(2) Conducted awareness-raising	56.7	86.8	96.1	+39.4*	+9.2
(3) Addressed exclusion	45.8	67.8	88.5	+42.7*	+20.7*
(4) Networked with CBOs/ institutions/other SBMCs	18.5	90.8	97.3	+78.8*	+6.5
(5) Interacted with LGEA	23.1	35.4	45.3	+22.2*	+9.9
(6) Has women's committee	27	68.1	86.1	+59.1*	+18.0*
(7) Has children's committee	14.6	57.7	82.7	+68.0*	+25.0*
(8) Contributed resources for school	48.8	71.4	84.3	+35.5*	+12.9
(9) Chair visited school three or more times	27.2	35.4	44.7	+17.5	+9.3
School meets functioning SBMC standard (%)	29.6	75.1	95.5	+65.9*	+20.5*
Number of SBMC functionality criteria met (/9)	3.0	5.9	+4.0*	+1.2*	+4.0*
Additional indicators: inclusion and drop-out					
Took action for commonly excluded groups (%)	37.5	16.4	36.1	-1.4	+19.7*
Raised issue of children's exclusion (%)	4.9	13.3	23.2	+18.3*	+9.8
Additional indicators: Organising and mobilising resources					
School has an SBMC (%)	93.6	100	100	+6.4	+0.0
Cashbook available (%)	33	53.4	82	+49.0*	+28.6*
Requested support from LGEA or SUBEB (%)	n/a	n/a	87	n/a	n/a
Raised cash to support school improvement (%)	23.2	42.5	48.1	+24.9*	+5.6
Mobilised non-cash resources (%)	43.1	59.9	78.6	+35.6*	+18.7
Involved in making SDP (%)	63.5	74.8	n/a	+11.4	

* Indicates that change over time is statistically significant ($p < 0.5$)

Box 8: Asking SBMCs about inclusion and exclusion

A number of different criteria aim to measure the SBMC's inclusiveness and the actions it has taken on excluded children. These were based on the following questions addressed to the SBMC chairperson. As elsewhere, questions were asked in the local language, with instructions to use a language that the respondent could understand but not to provide additional explanation or prompts.

Criterion	**Question asked** (with data collector instructions in blue)	Criterion met if
(2) Conducted awareness-raising	Did the SBMC do anything to raise awareness about the value of education for all boys and girls in the community in the current school year?	Respondent answers yes and can present oral or written evidence
(3) Addressed exclusion	Did the SBMC do anything to address issues which prevent children from attending school or which cause drop-out in the current school year?	Respondent answers yes and can present oral or written evidence
(A1) Took action for commonly excluded groups	Did the SBMC do anything to **support commonly excluded groups** in the **current school year**? You can explain that commonly excluded groups could be orphans, nomadic children, girls, children with disability, ethnic or religious minorities, etc.	Respondent answers yes and can present oral or written evidence
(A2) Raised issues of children's exclusion	Did the SBMC **raise issues of children's exclusion** from school in the community, with the LGEA, or with the state government, in the **current school year**?	Respondent answers yes and can present oral or written evidence
(A3) Raised cash to support vulnerable children	Did the SBMC mobilise any cash to support vulnerable children in the current school year?	Respondent answers yes (no evidence required)
(A4) Monitored drop-out or non-attendance (A5) Communicated with school or community about drop-out (A6) Number of actions taken to address non-attendance	**What actions were taken** to address issues which prevent children from attending school or which cause drop-out in the **current school year**? Do not prompt. This is a multiple response question **– SELECT ALL THAT APPLY** • Monitoring drop-out • Monitoring non-attendance • Communicating with school about drop-out • Communicating with community about drop-out • Other (specify) • Don't know / refused	Respondent answers yes to a previous question (asking whether any action was taken to address these issues) and then provides this information in the follow-up question on what type of action and how many actions were taken. No specific evidence is required

Schools that have received Output 4 support are more likely to meet ESSPIN's logframe standard for a functioning SBMC (see Table 17). However, there are variations across indicators. For

instance, schools that have received such support are far more likely to show evidence that they have women's and children's committees, meet regularly, and contribute resources to schools. However, they are slightly less likely to take action for commonly excluded groups or raise issues of children's exclusion. As the table below highlights, for most indicators, the impact of one additional year of Output 4 support is fairly small and is not statistically significant.

Table 17: Kwara: SBMC functionality in CS3, by intervention group

Intervention group	None	Post-CS1	Pre-CS1	Estimated effect of one year of full intervention by 2016
(1) Two or more meetings this school year	68.2	83.1	90	4.8
(2) Conducted awareness-raising	100	93.5	94.9	-1.6
(3) Addressed exclusion	91.3	85.2	88	-1.4
(4) Networked with CBOs/institutions/other SBMCs	100	100	95.4	-0.5
(5) Interacted with LGEA	34.8	38.8	52	1.7
(6) Has women's committee	72	96.1	90.1	7.1
(7) Has children's committee	42.6	92.2	99.4	15.2*
(8) Contributed resources for school	67.5	89.8	90.9	6.6
(9) Chair visited school three or more times	26.1	41	54.7	4.8
School meeting functioning SBMC standard (%)	89.4	90.5	99.8	1.0
Number of SBMC functionality criteria met (/9)	6	7.1	7.6	0.3*
Additional criteria				
Action for commonly excluded groups	56.3	52.7	21.5	-2.8
Raised issue of children's exclusion	43	30	11.6	-4.6
Raised cash to support vulnerable children (%)	65	57.4	28.9	-4.0
Requested support from LGEA or SUBEB (%)	97.6	77.1	84.5	-4.8*
Monitored drop-out or non-attendance (%)	100	68.3	85.7	-7.0*
Communicated with school or community about drop-out (%)	89.4	77.3	80.1	-2.0
No. of actions taken to address non-attendance	2.5	2.2	2.1	-0.1
* indicates that estimated effect of one year of full intervention is statistically significant ($p < .05$)				

3.3.3 How inclusive are SBMCs of women and children?

As noted above, SBMCs are expected to have women's and children's committees. We also record a number of other measures of the extent to which SBMCs are inclusive of women's and children's concerns. In each case, there are four criteria and an overall standard (Box 9).

Box 9: Logframe criteria for SBMCs' inclusiveness of women and children

The school must meet at least three of the four criteria listed below in order to meet the SBMC **women's inclusiveness** standard for the last school year:

1) at least one woman attended two or more SBMC meetings (written evidence);
2) female member of SBMC raised at least one issue at SBMC meetings (written evidence or oral evidence from a female member of the SBMC);
3) at least one issue raised by a female member at an SBMC meeting led to action (written, physical or oral evidence from a female member of the SBMC); and
4) at least one SBMC women's committee meeting took place.[5]

The school must meet at least three of the four criteria listed below in order to meet the SBMC's **children's inclusiveness** standard for the current school year:

1) at least one child attended two or more SBMC meetings (written evidence);
2) a child member of SBMC raised at least one issue at SBMC meetings (written evidence or oral evidence from child member of SBMC);
3) at least one issue raised by a child member at an SBMC meeting led to action (written, physical or oral evidence from child member of SBMC); and
4) at least one SBMC children's committee meeting took place and the committee has a trained facilitator.[6]

CS2 recorded a sharp increase in the share of SBMCs meeting ESSPIN's standard for women's inclusiveness. This was driven by large improvements in all four criteria associated with this indicator. All four criteria continued to improve between 2014 and 2016, although the extent of these improvements was more modest, which is unsurprising given that they started from a higher base. In 2016, 78% of SBMCs met the standard for women's inclusiveness.

Table 18: Kwara: SBMC women's inclusiveness in CS1, CS2 and CS3

	2012 (CS1)	2014 (CS2)	2016 (CS3)	Change: 2012 vs. 2016	Change: 2014 vs. 2016
(1) At least one woman attended two or more meetings (%)	37.1	58.5	65.2	+28.0*	+6.7
(2) Female member raised an issue (%)	44.2	84.8	89.5	+45.3*	+4.7
(3) Issue raised by female member led to action (%)	44.9	65.2	76.1	+31.2*	+10.9
(4) Women's committee met (%)	5.2	73.6	81.3	+76.1*	+7.7
Number of criteria met (/4)	1.3	2.8	3.1	+1.9*	+0.3
Meets women's inclusiveness standard (3/4) (%)	26.6	70.8	78.2	+51.6*	+7.4

* indicates that change over time is statistically significant ($p < .05$)

Schools that have received Output 4 interventions were more likely to meet the women's inclusiveness standard in 2016 than those have not. CS3's findings suggest that one year of Output 4 intervention is associated with a 7.4 percentage point higher likelihood of meeting the standard. In

[5] This criterion has been slightly altered since CS1; in CS1 it also required the women's committee to have a female leader.
[6] In CS1 this criterion required written evidence in the form of minutes of at least one children's committee meeting held in the past school year. This requirement was dropped for CS2 as it was considered unlikely that children's committees would keep good minutes, and that a failure to keep minutes does not mean the committee is not functioning.

particular, schools that have received Output 4 support were significantly more likely to show evidence that an issue raised by a female member of the SBMC had led to action.

Table 19: Kwara: SBMC women's inclusiveness in CS3, by intervention group

Intervention group	None	Post-CS1	Pre-CS1	Estimated effect of one year of full intervention by 2016
(1) At least one woman attended two or more meetings (%)	47.4	79.2	69.8	5.9
(2) Female member raised an issue (%)	82.7	98.6	90.3	3.2
(3) Issue raised by female member led to action (%)	56	89.7	82	7.5*
(4) Women's committee met (%)	72	92.8	82.5	4.7
Number of criteria met (/4)	2.6	3.6	3.2	0.2*
Meets women's inclusiveness standard (3/4) (%)	56	88.4	86.1	7.4*

* indicates that estimated effect of one year of full intervention is statistically significant ($p < .05$)

Following a significant improvement in the share of SBMCs meeting the children's inclusiveness standard between CS1 and CS2, there was no improvement in this indicator between CS2 and CS3. However, improvements continued to be recorded in some of the criteria for children's inclusiveness, notably the share of children's committees that had met and that have a trained facilitator. In 2016, only 26% of SBMCs met the children's inclusiveness standard. In particular, children's attendance at SBMC meetings remains poor (criteria 1). These findings raise some questions about the extent to which SBMCs are able to serve as a forum for children to raise concerns about the school and influence the school improvement agenda.

Table 20: Kwara: SBMC children's inclusiveness in CS1, CS2 and CS3

	2012 (CS1)	2014 (CS2)	2016 (CS3)	Change: 2012 vs. 2016	Change: 2014 vs. 2016
(1) At least one child attended two or more meetings (%)	25.1	24.5	26.5	+1.4	+2.0
(2) Child raised an issue (%)	27	56.3	54.2	+27.2*	-2.1
(3) Issue raised by child led to action (%)	20.7	31.5	42.5	+21.7*	+11.0
(4) Children's committee met and has a trained facilitator (%)	5.5	34.3	60.4	+54.9*	+26.1*
Number of criteria met (/4)	0.7	1.5	1.6	+0.9*	+0.2
Meets children's inclusiveness standard (3/4) (%)	11.1	26.6	26.4	+15.4*	-0.2

* indicates that change over time is statistically significant ($p < .05$)

Schools that have received some Output 4 support were more likely to meet the children's inclusiveness standard in 2016 than those that had not. However, schools that had received more years of intervention (pre-CS1 schools) were *less* likely to meet the standard than those that had received fewer years of intervention (post-CS1 schools).

Table 21: Kwara: SBMC children's inclusiveness in CS3, by intervention group

Intervention group	None	Post-CS1	Pre-CS1	Estimated effect of one year of full intervention by 2016
(1) At least one child attended two or more meetings (%)	29.5	47.9	19.3	1.0
(2) Child raised an issue (%)	44.4	73.5	52.5	5.2
(3) Issue raised by child led to action (%)	41.6	59.3	38.1	2.7
(4) Children's committee met and has a trained facilitator (%)	42.9	83.8	63.2	7.9*
Number of criteria met (/4)	1.4	2.4	1.5	0.1
Meets children's inclusiveness standard (3/4) (%)	19.8	56.8	21.7	5.7
* indicates that estimated effect of one year of full intervention is statistically significant ($p < .05$)				

4 Teachers

The SIP includes training for teachers on basic literacy and numeracy, and teaching skills (the use of teaching aids, participation and praise, and techniques for classroom organisation). This chapter first looks at changes in teacher competence as measured by ESSPIN's logframe indicator. It then looks in more detail at teachers' performance on a set of literacy and numeracy content knowledge tests. Finally it looks at changes in teacher motivation in Kwara.

4.1 Teacher competence

Box 10: Teacher competence: Key findings

- Following a small decline between 2012 and 2014, there has been no significant change in the share of teachers meeting ESSPIN's competence standard between 2014 and 2016. This has started from a high initial base – 85% of teachers met the competence standard in 2012, which is likely to partly reflect the early years of the SIP in Kwara.
- In 2016, 76% of teachers met ESSPIN's basic competence standard while only 30% met the strict version of this standard. The latter reflects the finding that only 39% of teachers were able to score at least 50% on primary school-level English literacy and numeracy tests.
- Teachers' performance on both the English literacy and numeracy tests declined significantly between CS2 and CS3.
- ESSPIN-trained teachers are more likely to meet the competence standard than non-ESSPIN-trained teachers and they also performed better than their peers on the English literacy and numeracy tests.

The ESSPIN logframe sets four criteria for judging the competence of teachers: one relates to curriculum knowledge (although this applies only to teachers who teach English or maths), two relate to teaching practices, and one relates to classroom organisation. Teachers are defined as competent if they meet three of the four criteria (two of the three relevant criteria in the case of those who do not teach English or maths; see Box 10).

For CS2 and CS3, a stricter version of the competence indicator was developed. The criterion of using at least one teaching aid during the lesson observation was changed to exclude reading from, writing on or having pupils copy from the blackboard, since this is considered to be a poor use of a teaching aid that is less likely to enhance learning. In addition, a fifth criterion was added that is based on teachers' performance on content knowledge tests. Teachers are defined as competent if they are competent according to the original criteria and can also score at least 50% on primary school-level literacy and numeracy tests.

Box 11: Criteria for teacher competence

A teacher must meet three out of four of the following criteria to meet the competence standard if he/she teaches English and/or maths. Teachers of other subjects must meet two out of three criteria (excluding 1 below):

1) knowledge of English or mathematics curriculum (based on interview);
2) use of at least one teaching aid during lesson observation;
3) greater use of praise than reprimands during lesson observation; and
4) in terms of class organisation: assigning individual or group tasks at least twice during lesson observation (or for two contiguous five-minute blocks).

For CS2 and CS3, stricter criteria for teacher competence were introduced. These modified (2) to exclude reading from or writing on, or having pupils copy from, the blackboard as a use of a teaching aid. A fifth criterion was added:

5) literacy and numeracy: scores at least 50% in both an English literacy and a numeracy test.

Following a small decline between 2012 and 2014, there has been no significant change in the share of teachers meeting ESSPIN's competence standard between 2014 and 2016. This has started from a high initial base – 85% of teachers met the competence standard in 2012. This is likely to partly reflect the early years of the SIP in Kwara. The share of teachers that meet the strict version of the competence standard fell between 2014 and 2016, although this change was not statistically significant. Overall, in 2016 76% of teachers met ESSPIN's basic competence standard, while only 30% met the strict version of this standard. The latter indicator reflects the striking finding that only 39% of teachers were able to score at least 50% on primary school-level English literacy and numeracy tests.

When interpreting findings related to the basic competence standard, it is worth noting that there are some concerns about the accuracy of the teacher competence estimates in CS1. This is because lesson observations were done on paper rather than on CAPI (which restricted the scope for checking their accuracy) and because of the way that questions on the curriculum benchmarks were administered during CS1. As a result, it is possible that teacher competence may have been over-estimated in CS1.

Table 22: Kwara: Teacher competence in CS1, CS2 and CS3

	2012 (CS1)	2014 (CS2)	2016 (CS3)	Change: 2012 vs. 2016	Change: 2014 vs. 2016
(1) Knowledge of Eng/maths curriculum [1]	79.0	52.3	56.9	-22.1***	+4.6
(2) Use of one or more teaching aid	89.4	97.1	99.8	+10.3**	+2.7
(2a) Use of one or more teaching aid, excl. blackboard	0.0	84.2	89.8	n/a	+5.6*
(3) Praise more than reprimand	83.9	91.8	95.6	+11.7*	+3.8
(4) Assigns two or more ind./group tasks	61.1	58.2	45.5	-15.7*	-12.7**
(5) Passes English and maths test	n/a	45.0	38.6	n/a	-6.4
Teacher competence score (% of criteria met)	77.4	75.0	74.6	-2.8	-0.5
Teacher competence standard fulfilled	84.7	73.6	76.3	-8.4	+2.7
Teacher competence score (% of criteria met, strict)		70.1	68.5	n/a	-1.6
Teacher competence std. fulfilled (strict; %)		38.4	30.1	n/a	-8.3

* indicates that change over time is statistically significant ($p < 0.5$)

1. There were some inconsistencies in how questions related to curriculum knowledge were administered in CS1 compared to other rounds of the survey. This may partly explain the sharp drop in curriculum knowledge between CS1 and CS2.[7]

The detailed findings in Table 22 indicate that, as at 2016, the vast majority of teachers are using at least one teaching aid and making greater use of praise than reprimands. However, the propensity to assign individual and group tasks is low and there are persisting gaps in the curriculum knowledge of many English and maths teachers. Furthermore, there are large gaps in teachers' subject knowledge, and this has worsened slightly between 2014 and 2016. It would be worth exploring why the SIP has not made greater progress on these fronts, and what could be done to address this going forward.

We also examine how teachers who report having been trained by ESSPIN by CS3 perform compared to those that do not. When interpreting the findings below it is worth noting, however, that it is part of ESSPIN's model that trained teachers should share their training with the rest of the school staff, with support from head teachers. As a result, some spillover effects are expected, and the difference between ESSPIN-trained and non-ESSPIN-trained teachers cannot be interpreted as a clear indicator of ESSPIN's impact. A simple comparison of the two would underestimate ESSPIN's impact on teacher competence, to the extent that it masks the positive diffusion effect on non-directly trained teachers.

The findings below indicate that ESSPIN-trained teachers perform better than non-ESSPIN-trained teachers on every indicator of teacher competence, although in some cases this difference is not statistically significant. In particular, ESSPIN-trained teachers are much more likely to have adequate curriculum knowledge and to assign individual and group tasks to pupils. Overall, 85% of ESSPIN-trained teachers met the teacher competence standard in 2016, compared to 62% of non-

[7] CS2 introduced clearer guidance about which grade of the curriculum teachers should be quizzed on, in order to improve consistency within the CS2 data. In addition, CS1 fieldwork in each school was spread over several days, giving teachers an opportunity to revise their knowledge of curricula guidelines. In CS2, fieldwork in each school was conducted on a single day.

ESSPIN-trained teachers. ESSPIN-trained teachers were also more likely to meet the stricter competence standard, although this difference is not statistically significant.

Table 23: Teacher competence in CS3 by intervention group

Intervention group	Non-ESSPIN-trained	ESSPIN-trained	Difference
(1) Knowledge of Eng/maths curriculum (%)	45.0	64.4	+19.5*
(2a) Use of one or more teaching aid (%)	99.6	99.9	+0.3
(2b) Use of one or more teaching aid, excl. blackboard (%)	85.1	93.2	+8.1
(3) Praise more than reprimand (%)	94.0	96.9	+2.9
(4) Assigns two or more ind./group tasks (%)	38.2	50.3	+12.0
(5) Passes English and mathematics test	33.7	41.7	+8.0
Teacher competence score (% of criteria fulfilled)	69.6	77.9	+8.3*
Teacher competence standard fulfilled (%)	62.3	85.4	+23.1*
Teacher competence score (% of criteria fulfilled, strict version)	62.8	72.3	+9.5*
Teacher competence standard fulfilled (strict) (%)	24.4	33.7	+9.3
Additional indicators:			
Proportion of time spent – explaining (%)	45.8	44.6	-1.2
– instructing / presenting / dictating (%)	24.3	24.7	+0.4
– chanting (%)	4.6	4.0	-0.6
– asking closed question / response (%)	3.9	3.6	-0.3
– asking open question / response (%)	2.7	3.5	+0.8
Proportion of time spent speaking English (%)	67.7	70.5	+2.7
Teacher summarised the lesson (%)	76.8	79.1	+2.3
Teacher revisited the lesson's objectives (%)	48.5	44.3	-4.2
Teacher gave learners homework (%)	30.1	26.9	-3.2
Teacher tested learners' knowledge (%)	66.0	75.4	+9.4
Teacher marked learners' written work (%)	40.9	49.0	+8.0
* indicates that change over time is statistically significant (p < 0.5)			

CS3 also collected additional information on what teachers do during their lessons. The findings indicate that there is not much difference in the method of teaching that is used by ESSPIN-trained and non-ESSPIN-trained teachers. Both spent similar amounts of time explaining, instructing, chanting, and asking closed and open questions. However, ESSPIN-trained teachers were more likely to test learners' knowledge and mark learners' written work.

4.1.1 Teacher content knowledge tests

Approach to analysing the results of the teacher tests

Percentage scores in the teacher content knowledge tests provide a rough indication of teachers' test performance, but analysis using item response theory (IRT) provides more reliable learning scales that can also be interpreted more readily in terms of learning benchmarks (see Allen, 2016a). Teachers' results can be divided into four performance bands in literacy and five performance bands in numeracy. Review of the items that teachers in each band can mostly

answer correctly then provides descriptors for each band (Table 24). For example, a teacher in Band 2 for literacy is one who shows knowledge of some basic phonics, can write a simple sentence, and can demonstrate basic comprehension of a passage, as well as being able to satisfy the easier items – testing limited comprehension of simple passages, basic nouns and verbs – associated with a teacher in Band 1. The teacher in Band 2 cannot typically correctly answer the harder items associated with Bands 3 or 4, such as identifying simple antonyms.

Table 24: Band descriptors based on IRT analysis

Band	Literacy	Numeracy
5		Understands conversion of fractions to decimals, and place values in decimals
4	Creates several sentences, shows knowledge of phonics, punctuation, formal letter layout, suffixes and alphabetical order	Understands ideas of area, nets, pictograms and rounding
3	Past/present of verbs, completes a sentence, extracts basic information from a passage, identifies simple antonyms, forms plurals	Understands basic sets, use of the number line to represent sums, conversion of units of time and mass, can complete word problems involving division
2	Shows knowledge of some basic phonics, writes a simple sentence, basic comprehension of a passage	Simple division, word problems involving addition, signs for arithmetic operations, integer comparisons and integer place values
1	Limited comprehension of simple passages, basic nouns and verbs	Simple addition with carrying over, simple subtraction, identifying a fraction, counting, simple regular shapes

Within the literacy and numeracy tests, items can be grouped according to specific sub-domains of learning: reading, writing and grammar within literacy, and number concepts and calculation within numeracy.

Findings

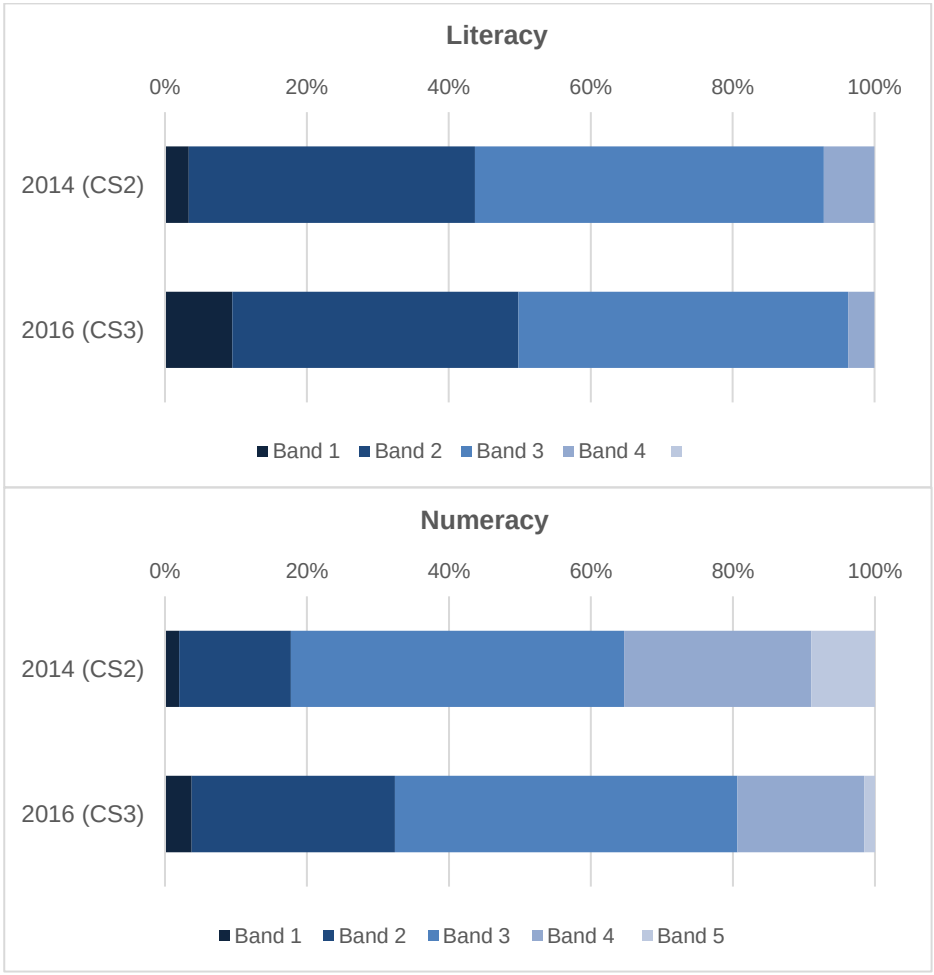
The CS3 findings point to a significant deterioration in teachers' performance on both the English literacy and numeracy tests. Teachers' scale scores in English declined by 0.2 standard deviations and those in mathematics fell by 0.4 standard deviations (see Table 26). In line with this, the share of teachers in the lower two bands increased from 44% to 50% in the case of English and from 18% to 32% in the case of mathematics.

Table 25: Teachers' test scores (IRT analysis) in 2014 and 2016 in Kwara

	2014 (CS2)	2016 (CS3)	Change in average, 2014–16
English IRT scale score (mean 500, s.d. 100)	506.8	486.6	**-20.1***
English Band 1 (%)	3.4	9.5	**+6.1***
English Band 2 (%)	40.3	40.3	+0.0
English Band 3 (%)	49.2	46.5	-2.7
English Band 4 (%)	7.1	3.7	-3.4
Reading (English sub-scale, mean 500, s.d. 100)	515.2	493.0	**-22.2***

Writing (English sub-scale, mean 500, s.d. 100)	483.6	473.6	-9.9
Grammar (maths sub-scale, mean 500, s.d. 100)	509.8	487.2	**-22.7***
Mathematics IRT scale score (mean 500, s.d. 100)	533.7	493.1	**-40.6***
Mathematics Band 1 (%)	2.0	3.7	+1.8
Mathematics Band 2 (%)	15.8	28.6	**+12.9***
Mathematics Band 3 (%)	46.9	48.2	+1.3
Mathematics Band 4 (%)	26.4	17.9	**-8.5***
Mathematics Band 5 (%)	8.9	1.5	**-7.4***
Number concepts (maths sub-scale, mean 500, s.d. 100)	532.4	491.4	**-41.1***
Calculation (maths sub-scale, mean 500, s.d. 100)	532.3	497.2	**-35.1***
* Indicates that change over time is statistically significant (p < 0.5)			

Figure 2: Share of teachers in each performance band in 2014 and 2016

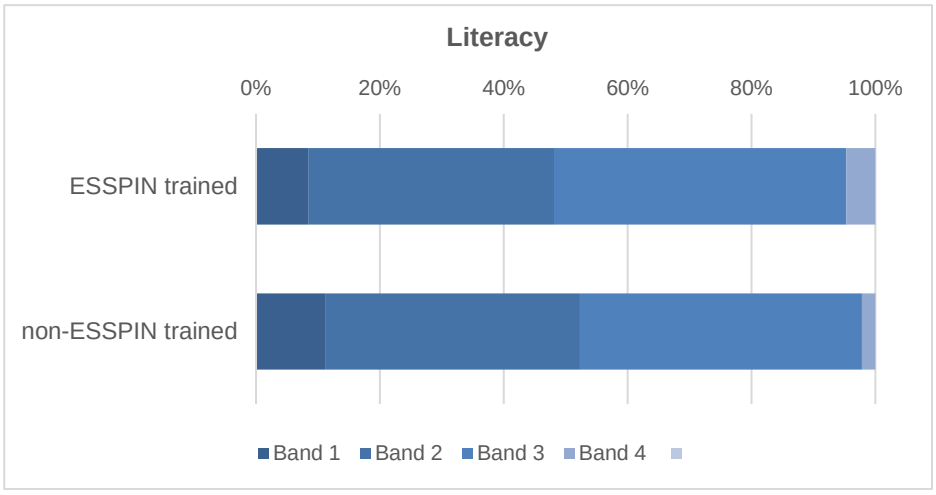


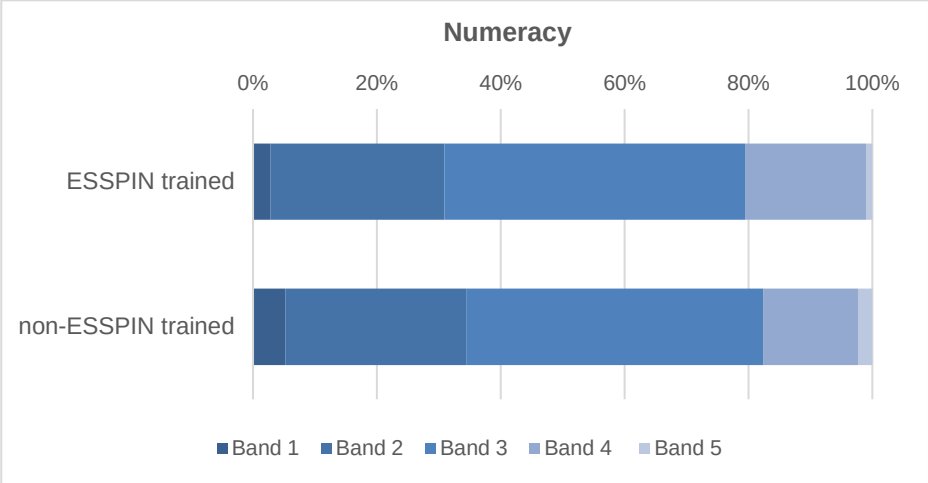
Teachers who received ESSPIN training performed better than those who did not, in both English and mathematics, although these differences were not statistically significant. In both cases, there is a difference of about 0.1 standard deviations between scale scores for ESSPIN-trained and non-ESSPIN-trained teachers. As mentioned above, the implications of this difference are not entirely straightforward given that ESSPIN anticipates that teachers directly trained as part of the SIP will share their knowledge and skills with other teachers in their school.

Table 26: Teachers' test scores (IRT analysis) by ESSPIN training in Kwara

	Non-ESSPIN-trained	ESSPIN-trained	Difference in means
English IRT scale score (mean 500, s.d. 100)	484.2	497.3	+13.1
English Band 1 (%)	10.0	7.6	-2.4
English Band 2 (%)	37.8	43.9	+6.1
English Band 3 (%)	50.2	43.0	-7.2
English Band 4 (%)	1.9	5.5	+3.6
Reading (English sub-scale, mean 500, s.d. 100)	493.7	501.7	+8.0
Writing (English sub-scale, mean 500, s.d. 100)	467.7	483.0	+15.3
Grammar (maths sub-scale, mean 500, s.d. 100)	484.0	499.3	+15.4
Mathematics IRT scale score (mean 500, s.d. 100)	494.3	505.0	+10.7
Mathematics Band 1 (%)	0.6	5.0	+4.5
Mathematics Band 2 (%)	31.9	21.6	-10.2
Mathematics Band 3 (%)	53.0	45.4	-7.6
Mathematics Band 4 (%)	14.4	24.7	+10.3
Mathematics Band 5 (%)	0.2	3.3	+3.0
Number concepts (maths sub-scale, mean 500, s.d. 100)	491.2	505.1	+14.0
Calculation (maths sub-scale, mean 500, s.d. 100)	500.0	505.0	+5.0
* Indicates that change over time is statistically significant (p < 0.5)			

Figure 3: Share of ESSPIN-trained and non-ESSPIN-trained teachers in each performance band in 2016





4.2 Teacher motivation

Teacher motivation may have a notable influence on the extent to which ESSPIN's contributions to teacher competence translate into improved teaching practices – if teachers are demotivated, they may be less likely to apply these skills or to attend school regularly. Teacher motivation may also be influenced by training and mentoring – as teachers acquire new skills, their self-efficacy may increase, in turn making them more engaged in, and committed to, their jobs.

Box 12: Measuring teacher motivation

For CS3, we included a measure of teacher motivation and teacher interaction using a scale that had been developed for the Nigerian context and used and tested in two previous school-based surveys. We define teacher motivation as the propensity of teachers to start and maintain behaviours that are directed towards fulfilling their professional goals, and in particular towards achieving better learning outcomes for the school's learners (Cameron, 2015b). Many existing instruments designed to measure teacher motivation focus exclusively on 'efficacy' – the extent to which teachers see themselves as able to influence their pupils' learning outcomes – which can also be seen as the 'can do' aspect of motivation (Bennell and Akyeampong, 2007). We wished to go beyond this to include measures relating more closely to teachers' willingness to work hard and their commitment, effort and enjoyment, which might together be labelled as 'will do' aspects of motivation.

The motivation scale we developed was incorporated into the teacher interview. Teachers were asked to what extent they agreed[8] with a series of statements that measure different aspects of motivation. The scale consists of three sub-scales of teacher motivation (satisfaction, skills and engagement) and one scale of teacher–teacher interaction (collegiality). The three sub-scales of teacher motivation were combined into a composite motivation measure by calculating the mean of the three sub-scales[9]. The teacher motivation scale was also analysed using IRT. Table 27 describes each of the different sub-scales and provides some examples of the items used to assess these.

Source: CS3 Technical Report.

[8] Teachers were asked to pick from amongst the following options: 'Strongly disagree', 'Disagree', 'Agree', 'Strongly agree'

[9] The three sub-scales were also combined into a composite measure using partially non-compensatory methods. These produced composite measures which were very highly correlated with the simple mean composite.

Table 27: Teacher motivation and interaction scale and sub-scales

Scale	Description	Example of items
Collegiality	How I see the extent of commitment and collaboration among my colleagues ('teacher–teacher interaction')	• All of the teachers in my school trust each other • All teachers at this school are highly committed to their job
Satisfaction	The value I place on my role as a teacher ('interest and enjoyment')	• I always enjoy teaching very much • I like to spend a lot of energy to make my classes interesting
Skills	The perception I have of my competences and skills as a teacher ('self-efficacy')	• I believe I know how to teach well • I believe I have the skills needed to encourage my learners to always work hard
Engagement	How engaged and committed I feel I am to my work as a teacher ('pressure/tension')	• It is difficult to manage learners in my classroom • Teaching is very tiring
Composite measure	Mean of satisfaction, skills and engagement sub-scales	

ESSPIN-trained teachers appear more motivated than non-ESSPIN-trained teachers, although this difference is not statistically significant (see Table 28). They appear to be more motivated than non-ESSPIN-trained teachers on three of the four sub-scales, particularly the skills sub-scale, which measures teachers' perceptions of their competence and skills. These findings suggest that training provided by ESSPIN has improved teachers' perception of their skills as well as other aspects of their motivation. However, it is also possible that some teachers were selected for training partly because they were more motivated or had other characteristics associated with motivation, such as higher qualifications. This would need to be investigated further.

Table 28: Teacher motivation and interaction by ESSPIN training

	Non-ESSPIN-trained	ESSPIN-trained	Difference in means
Collegiality	497.8	515.4	+17.6
Satisfaction	503.5	516.9	+13.4
Skills	487.0	517.5	**+30.6***
Engagement	539.9	539.9	+0.0
Composite motivation measure	509.4	524.2	+14.8

Note: All scores are normalised to have an average (mean) of 500 and a standard deviation of 100.

5 Trends in school quality

5.1 School quality

Box 13: School quality: Key findings

- The share of schools that meet ESSPIN's quality standard has improved significantly from 11% in 2012 to 38% in 2016. However, when we use a strict version of this standard, which incorporates teachers' performance on the literacy and numeracy tests, we find that schools' performance in Kwara has stagnated.
- Schools that have received ESSPIN's Output 4 interventions performed better on almost all measures of school quality in 2016.
- These schools also recorded significantly faster improvements in their quality scores between 2012 and 2016, suggesting that ESSPIN's Output 4 support has made a positive contribution to school quality in Kwara.

The ESSPIN logframe defines an overall measure of school quality that draws on the standards for teacher competence, head teacher effectiveness, school development planning, and SBMC functionality. A quality school is defined as one that meets the teacher competence standard and at least two of the other standards (see Box 14). We also use a 'quality score' indicator, which is an average of the scores that schools achieve on each of the four indicators mentioned above.

Box 14: Logframe standard for school quality

The school must meet at least three of the four output standards listed below in order to meet the school quality outcome standard, with teacher competence having to be one of those three.

1) teacher competence standard (more than half the teachers sampled in each school must be competent);
2) head teacher effectiveness standard;
3) school development planning effectiveness standard; and
4) SBMC functionality standard.

The version of this standard used in CS1 did not rely on teacher content knowledge tests. For CS2, we introduced a second, more strict version of the standard, in which teachers had to get above 50% in literacy and numeracy tests to be classed as competent (see Section 4.1 and Box 12 above).

5.1.1 Changes in state-level outcomes between CS1, CS2 and CS3

Overall school quality has improved significantly in Kwara, from 11% of schools meeting the quality standard in 2012 to 38% meeting it in 2016. This means that roughly 190 additional schools met the quality standard in 2016 compared to 2012. These schools teach around 30,500 students, who are now studying in a better-quality environment, as per this measure.

The share of schools that meet the strict version of the quality standard stagnated between 2014 and 2016 and remains low, at 16%. This reflects the low levels of subject knowledge in English and mathematics among teachers in Kwara (although teachers in the state do outperform those in Kano and Jigawa, and have similar levels of subject knowledge to those in Kaduna[10]).

[10] This comparison is based on IRT scale scores.

Table 29: Kwara: School quality in 2012–16

	2012 (CS1)	2014 (CS2)	2016 (CS3)	Change 2012–16	Change 2014–16
Quality score (%)	45.9	57.2	60	**+14.2***	+2.8
School meets quality standard (%)	11.4	26.8	37.7	**+26.3***	+11.0
Quality score (strict version) (%)		55.9	59.8	n/a	+4.0
School meets quality standard (CS2 version) (%)		17.8	16.2	n/a	-1.5
* indicates that change over time is statistically significant (p < .05)					

5.1.2 Performance of schools in different intervention groups in 2016

Schools that have received Output 4 interventions performed better on almost all measures of the quality score in 2016 (see Table 30). These differences are statistically significant in the case of the two quality scores, but not in the case of the share of schools meeting ESSPIN's quality standards. We estimate that one additional year of Output 4 intervention is associated with a 2.5 point increase in schools' quality scores.

Table 30: Kwara: School quality by Output 4 intervention group, 2016

	None	Post-CS1	Pre-CS1	Estimated effect of one year of full intervention
Quality score (%)	52.5	61.1	63.3	**2.5***
School meets quality standard (%)	13.4	29.8	51.6	6.2
Quality score (strict version) (%)	50.6	59.1	61.5	**2.6***
School meets quality standard (CS2 version) (%)	8.7	5.5	22.9	0.8
* indicates that estimated effect is statistically significant (p < .05)				

5.1.3 Rates of improvement: Variations across intervention groups

The superior performance of schools that have received Output 4 support may reflect differences at baseline, rather than the contribution of ESSPIN's interventions. To explore this, we compare changes over time between schools that did not receive Output 4 support and those that did. We find that schools that received two to four years of ESSPIN Output 4 interventions recorded significantly faster improvements in their quality score between 2012 and 2016. Schools that received no Output 4 support saw their quality score improve by 4.6 points, while those that did receive such support saw their scores improve by 17.1 percentage points, on average. These findings suggest that ESSPIN's Output 4 interventions have made a positive contribution to school quality in Kwara.

Table 31: Kwara: Change over time in the school quality score (2012–16) – Differences between intervention groups

	Output 4 intervention during 2011/12–2014/15		
	No intervention	**Two to four years**	**Difference**
2012 (CS1)	47.9	44.7	-3.2
2014 (CS2)	45	62.4	17.5
2016 (CS3)	52.5	61.8	9.3
Difference (2012–2016)	4.6	17.1	12.5 *

* indicates that difference is statistically significant (p < .05)

6 Learning outcomes

Box 15: Pupil learning outcomes: Key findings

- Grade 2 literacy and numeracy outcomes and Grade 4 literacy outcomes have deteriorated between 2012 and 2016 in Kwara. Grade 4 numeracy outcomes improved marginally over this period.
- On three of the four assessments, there were only minor differences in learning outcomes between schools that had received Output 4 support and those that had not. The exception was Grade 4 numeracy, on which the former set of schools performed significantly better than the latter in 2016.
- After controlling for a variety of school-level characteristics and test scores in 2012, we do not find evidence to suggest that ESSPIN's Output 4 interventions have made a positive contribution to learning outcomes in Kwara. However, this finding provides limited insights into ESSPIN's overall impact in the state as it does not incorporate the effects of ESSPIN's Output 3 interventions, which have been received by all schools in the state. To the extent that Output 4 is focused principally on equitable access and participation of marginalised children, it could also be argued that success in Output 4 activities could tend to lower average learning outcomes in public schools in the state.[11]

ESSPIN's final goal – according to the most recent iteration of its logframe – is to contribute to improvements in learning outcomes in government schools in the six states in which it works. This chapter examines trends in learning outcomes in Kwara between 2012 and 2016, and differences in these outcomes across schools that have received different levels of Output 4 support. It also seeks to assess the extent to which changes in learning outcomes in Kwara can be attributed to ESSPIN.

6.1 Approach to measuring learning outcomes

Learning outcomes were measured in literacy and numeracy at Grades 2 and 4, and analysed using IRT (see Allen, 2016b and Allen, 2016c). The analysis for each test produces a scale score which, by design, has an average value of 500 and a standard deviation of 100. This scale is also divided into bands, indicating the level of proficiency of the learner. For the Composite Surveys, bands have been designed to correspond to the levels of proficiency expected at each grade in the Nigerian curriculum. For example, a learner in Band 2 for literacy is one who is able to demonstrate knowledge and skills in at least some of the tasks that are considered to be within the range of Grade 2 proficiency. Table 32 and

[11] While Annual School Census data indicate that overall enrolment in public schools in Kwara has not increased much since 2009/10, it is possible that the profile of children in public schools has changed over this period, as better-off children have shifted to the expanding private school sector, while interventions to boost access (including those supported by ESSPIN) have contributed to more marginalised children entering the school system.

Table 33 list some examples of the tasks within each band.

Table 32: Examples of knowledge and skills that learners in each literacy band can demonstrate

Band	Associated knowledge and skills
Band 4: Grade 4 and above	Read and understand the grammatical structure of a sentence and complete a missing word using 'where', 'which', 'what' and 'who'. Follow the conventions of letter-writing to complete a letter template. Completing grammatically accurate sentences, with correct spelling, and a greeting and sign off. Independently read for meaning a short, simple text with a range of sentence structures.
Band 3: Grade 3 literacy	Read phonically decodable two-syllable and three-syllable words that include common diagraphs and adjacent consonants. Independently plan and write a grammatically correct simple sentence. Read a simple sentence for meaning and complete a missing word using correct spelling.
Band 2: Grade 2 literacy	Use phonic knowledge to utter initial sounds of familiar animal names. Use knowledge of common inflections in spellings, plurals, to write the answer to a question. Spell simple high frequency words accurately.
Band 1: Emerging literacy	Verbally compose a short grammatically correct sentence in the continuous present tense in response to a question about a picture. Listen to a short passage and remember specific details to respond verbally to a question. Clearly shaped and correctly orientated copying of words, with an understanding of space and full stops.
Band 0: Pre-literacy	Understand and respond verbally with a grammatically correct sentence to a simple question about their age. Understand and respond verbally with a grammatically correct sentence to a simple question about their name. Use phonic knowledge to utter initial sounds of names of familiar objects and animals.

Table 33: Examples of knowledge and skills that learners in each numeracy band can demonstrate

Band	Associated knowledge and skills
Band 5: Grade 5 & above	Solve a word problem involving differences in time. Determine which number rule was used to make one number into another. Solve a simple algebra problem.
Band 4: Grade 4 numeracy	Being able to gather information by interpreting simple graphs. Calculate the area of a rectangle, multiplying a decimal number, to one decimal place, by a one-digit number, and record the answer in m2. Choose the most appropriate strategy to subtract a decimal number, to two decimal places and a two-digit number, involving measure.
Band 3: Grade 3 numeracy	Multiply a two-digit number by a one-digit number. Use short division; subtract a two-digit number from a two-digit number crossing the tens boundary. Choose a strategy to add a three-digit number and a two-digit number crossing the tens boundary, involving money.
Band 2: Grade 2 numeracy	Use non-standard units of measure to compare the capacity of three containers. Subtract a two-digit number from a two-digit number. Name common 2D shapes. Extend counting past 800 and count in tens.
Band 1: Emerging numeracy	Recognise and complete a sequence of three two-digit numbers that are multiples of five. Subtract a one-digit number from a two-digit number 1–19. Read analogue clock to the hour.
Band 0: Pre-numeracy	Compare the length of two straight lines. Use non-standard units of measure to compare the capacity of three containers. Count to 10.

6.2 Pupil learning achievement in English literacy and numeracy

6.2.1 Changes in state-level outcomes between CS1, CS2 and CS3

Since 2014, learning outcomes have deteriorated significantly for three of the four assessments carried out as part of the Composite Surveys. Grade 2 and Grade 4 literacy scores declined significantly between 2014 and 2016, offsetting the small improvements seen between 2012 and 2014. Having worsened between 2012 and 2014, Grade 2 numeracy scores continued to deteriorate. The share of Grade 2 pupils achieving numeracy outcomes consistent with their grade almost halved, from 51% in 2012 to 26% in 2016. The one exception to this discouraging picture is Grade 4 numeracy outcomes, which improved marginally over this period.

Table 34: Kwara: Learning outcomes, 2012–16

	2012 (CS1)	2014 (CS2)	2016 (CS3)	Change: 2012 vs. 2016	Change: 2014 vs. 2016
Grade 2 literacy score	495	510	479.9	-15.1	**-30.0***
Band 0: Pre-school (%)	42.4	36.9	50.3	+7.9	+13.5
Band 1: Grade 1 (%)	34.7	31.3	33.9	-0.8	+2.5
Band 2: Grade 2 (%)	22.9	31.8	15.8	**-7.1***	**-16.0***
Grade 4 literacy score	482.2	496.2	474.2	-8.0	**-21.9***

Band 1: Grade 1 (%)	60.2	54.6	62.6	+2.4	+8.0*
Band 2: Grade 2 (%)	21.5	16.5	17.2	-4.2	+0.7
Band 3: Grade 3 (%)	5.5	9.1	6.7	+1.2	-2.4
Band 4: Grade 4 (%)	12.9	19.8	13.5	+0.6	-6.3
Grade 2 numeracy score	547.4	516.2	483.3	**-64.0***	**-32.8***
Band 0: Pre-school (%)	1.1	0.7	3.1	+2.0	+2.4
Band 1: Grade 1 (%)	48.4	71.1	70.9	+22.5*	-0.2
Band 2: Grade 2 (%)	50.5	28.2	26	-24.5*	-2.2
Grade 4 numeracy score	498.2	490.5	500.5	+2.3	+10.0
Band 1: Grade 1 (%)	9.4	11	6.8	-2.6	-4.2
Band 2: Grade 2 (%)	38.6	41.8	44.2	+5.7	+2.5
Band 3: Grade 3 (%)	25.8	25.8	22.5	-3.2	-3.3
Band 4: Grade 4 (%)	22.3	15.6	13.8	-8.5*	-1.8
Band 5: Grade 5 (%)	4	5.8	12.7	**+8.7***	**+6.9***
* indicates that change over time is statistically significant (p < .05)					

There are a few potential explanations for the deterioration in learning outcomes in Kwara. First, poor subject knowledge amongst teachers is likely to be a key constraint and it is notable that this has also worsened between 2014 and 2016. While other aspects of teacher competence are now deemed to be relatively high, it is questionable how much impact one can hope to see from improved teaching techniques if teachers' mastery of primary grade-level content is so limited. Second, teacher absenteeism appears to have increased sharply, which may be linked to recent delays in salary payments. A second possible explanation is that the profile of learners at public schools in the state may be changing. It is possible that access to education is expanding, leading to the enrolment of previously excluded pupils who are likely to be from more disadvantaged backgrounds. It is difficult to confirm whether this is in fact the case – Annual School Census data suggest that enrolment in public schools in the state has stagnated. However, there are questions about the reliability of these data. In addition, it is possible that better off pupils are shifting to private schools in the state, so a decline in public school enrolment may still be consistent with a situation in which the profile of learners is switching towards those from more disadvantaged backgrounds.

6.2.2 Performance of schools in different intervention groups during CS3

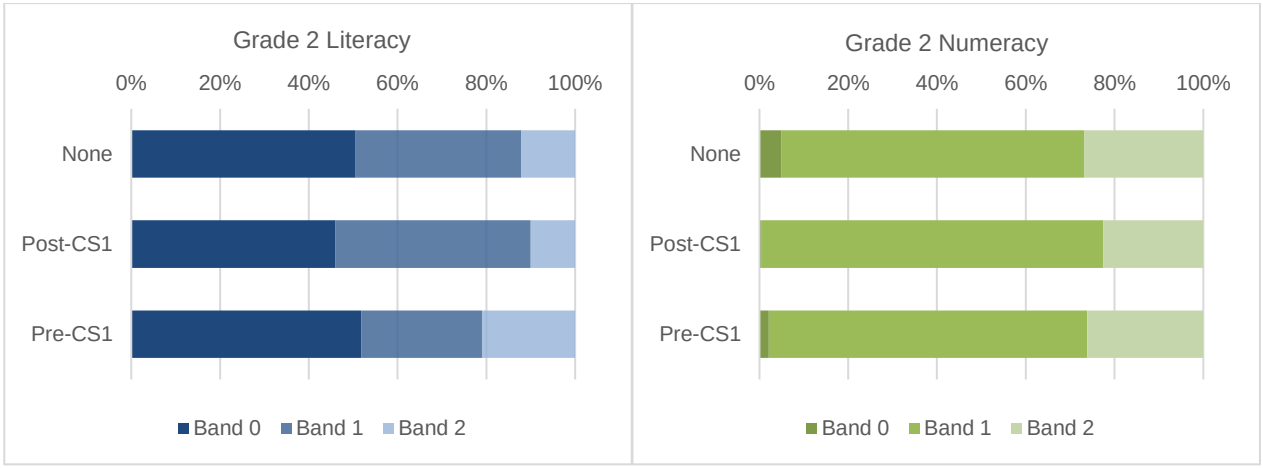
In 2016, for three of the four assessments there were only minor differences in learning outcomes between schools that had received Output 4 interventions and those that had not. However, for the Grade 4 numeracy tests, pupils in schools that had received Output 4 support performed significantly better than other schools in the state. This is reflected in Figure 4 below, which shows that in Output 4-supported schools, a significantly larger share of P4 pupils fall into the two highest proficiency bands for numeracy. It is worth reiterating that only limited inferences about ESSPIN's contribution to learning outcomes can be made based on these findings as they do not provide any insights into the role of ESSPIN support under Output 3.

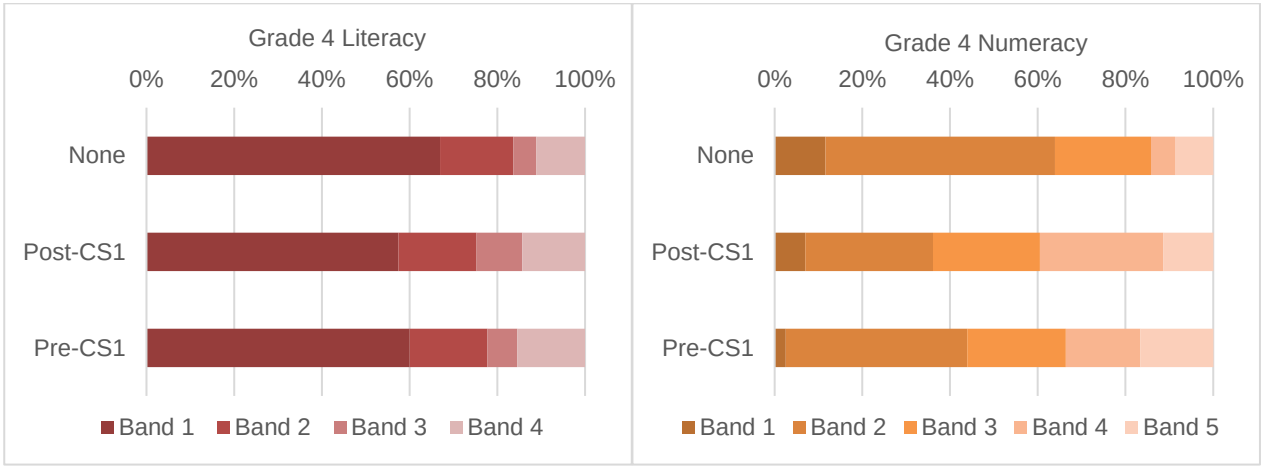
Table 35: Kwara: Learning outcomes by ESSPIN intervention group in 2016

	none	Post-CS1	Pre-CS1	Estimated effect of one year of full intervention

Grade 2 literacy score	474.3	477.8	485.7	0.5
Band 0: Pre-school (%)	50.5	46	51.8	0.3
Band 1: Grade 1 (%)	37.3	44.1	27.2	-0.2
Band 2: Grade 2 (%)	12.2	10	21	-0.1
Grade 4 literacy score	466	486.5	478.1	3.5
Band 1: Grade 1 (%)	67	57.4	60.1	-1.3
Band 2: Grade 2 (%)	16.7	17.8	17.7	-0.2
Band 3: Grade 3 (%)	5.3	10.4	6.8	0.6
Band 4: Grade 4 (%)	11.1	14.3	15.5	0.9
Grade 2 numeracy score	470.8	489.7	493.1	3.3
Band 0: Pre-school (%)	5	0.4	2.1	-1.0
Band 1: Grade 1 (%)	68.2	77.1	71.7	2.3
Band 2: Grade 2 (%)	26.9	22.5	26.1	-1.4
Grade 4 numeracy score	470.9	522.8	520.3	**11.5***
Band 1: Grade 1 (%)	11.6	7	2.5	-1.7
Band 2: Grade 2 (%)	52.3	29.1	41.4	-3.8
Band 3: Grade 3 (%)	21.9	24.4	22.5	0.2
Band 4: Grade 4 (%)	5.5	28.1	17	4.4*
Band 5: Grade 5 (%)	8.7	11.4	16.6	0.8
* indicates that estimated effect is statistically significant (p < .05)				

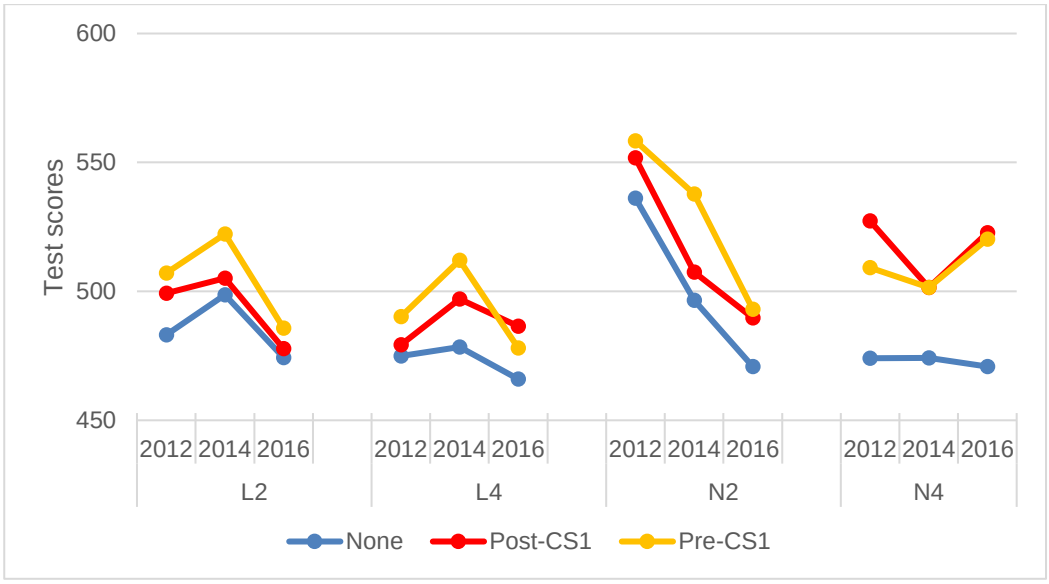
Figure 4: Kwara: Distribution of test scores by intervention group in 2016





In Figure 5, we disaggregate the change over time according to ESSPIN intervention. Schools that received Output 4 interventions had better learning outcomes at baseline in 2012 than schools that did not. Trends in learning outcomes seem to be fairly similar across the three intervention groups. These patterns are explored more rigorously in Sections 6.2 and 6.3 below, using regression and matching analysis to examine how change over time varies with ESSPIN intervention, and controlling for possible confounding variables, such as school characteristics.

Figure 5: Learning outcomes by test, year and intervention group



6.3 Controlling for school and pupil characteristics

6.3.1 Controlling for background characteristics

Schools in different intervention groups in Kwara differ from each other on several background characteristics. In particular, schools that have received no Output 4 interventions tend to be further away from LGA headquarters, and are more likely to be rural (see Annex A). They have slightly higher pupil–teacher ratios than schools that have received the most years of Output 4 support, although the difference is marginal. They are far less likely to have a power source than pre-CS1 schools (although not less than post-CS1 schools), and much less likely than post-CS1 schools to have classrooms with enough seating (although they are similar to pre-CS1 schools in this respect).

These differences could create an upward bias in our estimates of the effects of ESSPIN's Output 4 interventions on school-level outcomes. In particular, urban schools tend to have better learning outcomes than rural schools. We use a number of statistical methods to control for this type of 'confounding variable' – characteristics of schools that might affect learning outcomes and make it harder to tell whether the intervention is having an effect or not. We also estimate a model which controls for pre-existing differences in test scores by adding test scores in CS1 as a confounding variable. We use ordinary least squares regression analysis to estimate the models. Regression analysis estimates the correlation of learning outcomes with ESSPIN intervention, conditional on school characteristics.

After controlling for a variety of different school characteristics, we do find evidence of a statistically significant effect of a greater degree of ESSPIN Output 4 intervention on learning outcomes in Kwara for all learning outcomes apart from Grade 2 literacy (Model 2 in the table below). This indicates that schools with a greater degree of ESSPIN intervention had better learning outcomes in 2016, after controlling for a large set of confounding school characteristics.

Model 4 controls for both background characteristics and schools' test scores in 2012. The findings indicate that after controlling for background characteristics schools that received Output 4 interventions did not see faster improvement in their test scores between 2012 and 2014 than schools that did not receive such support. It is worth re-emphasising that these estimates do not give us an indication of ESSPIN's *overall* contribution to learning outcomes in Kwara because ESSPIN's Output 3 interventions have been received by all schools in the state. An additional caveat to note is that head teachers and SBMCs in schools that have not received Output 4 support may have also been exposed to ESSPIN's ideas and practices through informal communication or deliberate action by state of LGEA personnel.

Table 36: Estimates of the effect of ESSPIN Output 4 intervention on learning outcomes in 2016

Model	L2		L4		N2		N4	
(1) Simple regression, clustered SE, no sample weights	30.92	*	34.47	*	39.79	*	64.82	*
(2) Full covariates	21.62		23.01	*	44.46	*	63.86	*
(3) Lagged school-level learning outcomes	-1.97		0.96		-0.8		-1.25	
(4) Lagged outcomes and covariates	-3.09		0.12		1.01		-0.56	

7 Conclusions and implications of the Composite Surveys' findings for ESSPIN in Kwara

The Composite Surveys' findings paint a mixed picture with regards to the evolution of school-level outcomes in Kwara State. Since 2012, some indicators have improved significantly (notably SBMC functionality and school development planning effectiveness), some have levelled off (head teacher effectiveness), and others have worsened (school inclusiveness, teachers' competence and subject knowledge). Similarly, when one considers the level at which schools are performing as at 2016, outcomes are mixed. The vast majority of SBMCs in the state are functional, and almost 70% of schools are deemed to be inclusive. However, only 23% of head teachers are deemed to be effective and less than one-third of teachers meet ESSPIN's strict criteria for teacher competence.

Pupils' learning outcomes in English literacy and numeracy have largely deteriorated over this period. It is possible that this is partly related to the changing profile of learners in the state, although there is no clear evidence for this. Trends in learning outcomes may also be linked to the deterioration in teachers' subject knowledge and the rise in teacher absenteeism. ESSPIN has tried to compensate for gaps in teachers' content knowledge through the provision of scripted lesson plans to teachers. However, the Composite Survey findings raise potential questions about the extent to which such measures can contribute to improved learning in the face of fundamental weaknesses in teachers' own knowledge.

Contextual factors may partly explain the changes recorded by the Composite Surveys. In particular, the fiscal situation, and associated delays in teachers' salary payments, may well have undermined teacher motivation. This may be partly responsible for some of the findings related to teacher and head teacher effectiveness. It may also have repercussions in regard to the degree of attractiveness of the teaching profession, which is already low (and is arguably one of the reasons why the state struggles to attract people with high levels of subject knowledge into the teaching profession). Unlike most of the other ESSPIN states, however, pupil–teacher ratios in Kwara are low and have not increased significantly in recent years, suggesting that worsening learning outcomes are not linked to tighter resource constraints. The incidence of violence and insecurity is also relatively low in the state.

The implications of these findings for ESSPIN's contribution to school-level outcomes in Kwara are not entirely clear. Elements of the analysis that have sought to disentangle the effects of ESSPIN's interventions from broader trends in the state indicate that schools that have received Output 4 support have significantly higher SBMC functionality, more effective school development planning processes, and (as a result) higher school quality scores. These schools' quality scores have also improved more rapidly over time, suggesting that ESSPIN's support to SBMCs has made a positive contribution to school quality in Kwara. They also have better learning outcomes in 2016, after controlling for differences in background characteristics. However, they have not seen faster improvements in learning outcomes between 2012 and 2016 than schools that have not received Output interventions.

These findings are inherently limited in that they do not shed any light on the contributions of ESSPIN's Output 3 interventions, which have been implemented at scale since 2009. On balance, the findings on overall trends in the state suggest that these interventions have made positive contributions on some fronts, but had more muted results in others. As highlighted by a recent qualitative study on ESSPIN's capacity building interventions (Allsop *et al*. 2016), these findings are unsurprising given the difficult context in which ESSPIN works, notably the very low levels of subject knowledge amongst the primary-level teaching workforce.

The detailed findings discussed in the previous sections point to a number of key questions/issues for ESSPIN and its state-level partners. Some examples of these questions are given below.

- The decline in the share of head teachers taking action on teacher attendance, combined with the sharp increase in the teacher absenteeism rate, is notable. It is worth exploring what might be driving this, and particularly whether head teachers feel that they do not have the authority to influence teacher attendance, particularly in the context of delayed salary payments.
- A particularly striking finding from CS3 is that teachers have limited mastery of primary grade-level English and mathematics. Only 4% of teachers in our sample fall within the top proficiency band for English, and less than 2% fall into the top band for mathematics. It is questionable how much impact one can expect to see from better teaching skills if teachers' subject knowledge levels are low. This highlights the difficult environment in which ESSPIN's interventions have been implemented. It also underlines the importance of identifying what types of interventions should be prioritised in the face of such severe weaknesses in content knowledge.
- Although school development planning effectiveness has been improving in the state, the share of schools that implement four or more activities in their SDP remains low. Even amongst schools that have received Output 4 support, less than 30% met this criterion in 2016. The implementation of SDPs is clearly critical in order for more effective planning to be translated into improved school-level outcomes. As a result, weak performance on this front merits further investigation, particularly to assess whether this is something that is likely to improve with continued mentoring and support, or whether it is constrained by systemic factors (such as limited scope to raise funds from local communities, or for LGAs to respond to the school's needs).

Bibliography

Allsop, T., Nnodu, I., Jones, S., Rai, S. and Watts, M. (2016). Study of ESSPIN's Support to Capacity Development in Education in Nigeria. EDOREN.

Bennell, P., & Akyeampong, K. (2007). Teacher Motivation in Sub-Saharan Africa and South Asia (Department for International Development: Educational Papers). DFID.

Cameron, S. (2015b) Teacher motivation in the TDP baseline survey. Unpublished technical note. Oxford Policy Management.

ESSPIN (2009) School Based Management Committees in Policy and Practice, Volume 1: Research Synthesis Report. Report number ESSPIN 404.

ESSPIN (2013a) Overall findings and technical report of ESSPIN Composite Survey 1 (2012). Report number ESSPIN 060.

ESSPIN (2013c) Extension of the Education Sector Support Programme in Nigeria, August 2014–January 2017. Business case for DFID.

ESSPIN (2015) Kwara State report of ESSPIN Composite Survey 2 (2014).

Megill, D.J. (2014) Final Sample Design and Weighting Procedures for the ESSPIN Second Composite Survey (CS2) in Nigeria. Unpublished note written for ESSPIN and Oxford Policy Management.

Sanni, K. (2015) Taking School Improvement to Scale: The Education Sector Support Programme in Nigeria. Cambridge Education.

Annex A School background characteristics

The table below sets out summary statistics for Kwara's schools, split by categories according to the level of Output 4 intervention (none, post-CS1, pre-CS1). The data come from the Annual School Censuses for 2009/10, 2013/14 and 2014/15.

	Total	None	Post-CS1	Pre-CS1	
Distance from local government authority headquarters	**42.21**	50.67	34.94	16.23	*
Age of the school in 2014	**31.3**	32.75	27.05	37.92	
Urban (%)	**14.78**	6.68	23.08	34.67	*
Nomadic (%)	**2.22**	2.16	2.56	1.33	
Islamic (%)	**0.12**	0.22	0	0	-
Double shift (%)	**0.25**	0.43	0	0	-
Had parent–teacher association in 2014/15 (%)	**98.76**	98.7	99.63	96	
Had SBMC in 2014/15 (%)	**98.03**	96.77	100	98.67	*
PTR in 2009/10	**15.7**	20.52	12.34	13.78	*
PTR in 2013/14	**17.33**	20.11	14.93	17.07	*
PTR in 2014/15	**16.37**	16.25	17.55	12.47	
% in PTR between 2009/10 and 2013/14	**37.66**	39.32	32.79	45.53	
% in PTR between 2013/14 and 2014/15	**17.73**	18.16	15.54	23.17	
Number of classrooms in 2014/15	**5.57**	5.14	5.52	8.36	*
Number of teachers in 2014/15	**9.11**	7.45	9.91	16.47	*
Primary enrolment in 2009/10	**137.29**	123.34	150.8	133	*
Primary enrolment in 2013/14	**132.06**	111.45	145.97	142.63	*
Primary enrolment in 2014/15	**118.77**	97.14	131.86	205.21	*
% change in enrolment 2009/10–2014/15 (%)	**3.23**	-1.52	11.2	2.42	*
% change in enrolment 2013/14–2014/15	**7.42**	9.36	4.28	6.89	
% of teachers with academic diploma/degree	**69.05**	69.1	69.22	68.13	
% of teachers with PGDE, BEd or MEd	**16.53**	16.76	15.11	20.23	*
% of teachers with NCE, Grade II or equivalent	**71.77**	70.12	75.09	69.94	*
School has a power source (grid/other)	**9.96**	7.53	6.23	38.67	*
% of classrooms with enough seating	**45.84**	34.7	67.88	35	*
% of classrooms with a good blackboard	**75.65**	74.68	72.74	86.62	
% of classrooms in good condition/minor repairs	**44.68**	45.89	42.94	44.21	*
School has at least one toilet (%)	**68.09**	65.21	67.72	79.87	*
Number of schools	**819**	465	274	80	

Notes: (1) * indicates a significant coefficient when running a linear or logistic regression of the variable of interest (dependent variable) on the number of years of ESSPIN intervention (independent variable); (2) the column 'total' includes schools that do not have an intervention code.

Annex B ESSPIN Output 3 Interventions

The table below shows the ESSPIN Output Stream 3 interventions delivered to date in Kwara State. Each combination of interventions was categorised as ‘none’, ‘minimum’, ‘medium’, or ‘maximum’, according to the number of years of continuous intervention and hence expected impact.

Expected impact	Number of schools	2009/10			2010/11			2011/12			CS1	2012/13			2013/14			CS2	2014/15			2015/16			CS3
		L	T	SV	L	T	SV	L	T	SV		L	T	SV	L	T	SV		L	T	SV	L	T	SV	
Maximum (4)	1,485	6	3	30	6	3	30			30				30	6	3	30		6	3	30	2	4	15	

Note: L = days of leadership training; T = days of teaching training; SV = school visits.

Annex C ESSPIN Output 4 Interventions

The table below shows the days of Output 4 intervention in Kwara under different headings: SBMC training; women's and children's participation training; and mentoring visits.

Level of Output Stream 4 intervention	Number of schools	2010/11			2011/12			CS1	2012/13			2013/14			CS2	2014/15			2015/16			CS3
		S	P	M	S	P	M		S	P	M	S	P	M		S	P	M	S	P	M	
None	571																					
Pre-CS1	263				7		4		1		4		6	4		5	5	16	1	2	17	
Post-CS1	418											4		2		5	5	6	1	2	7	
Post-CS1	226											7		1		5	5	5	1	2	6	

Note: S = SBMC training; P = women and children participation training; M = mentoring visits; r = one-day refresher; mentoring visits were by civil society–government partnership teams, except those marked with an asterisk, which were by SMOs.

Pre-CS1 schools: those that received at least five days of any Output 4 intervention during or prior to 2011/12. Post-CS1 schools: those that have only received Output 4 interventions in 2012/13 or thereafter.

Annex D Difference-in-difference analysis using regressions

Test	treatment variable	model	coefficient	SE	P value	N	R-squared
L2	pu_exposure	Simple model with survey weights	-0.48	2.21	0.829	381	0.000332
L2	pu_exposure	No survey weights but clustered SEs	11.56	1.43	0	2836	0.064225
L2	intervention_binary	Binary exposure variable	30.92	15.01	0.042	381	0.024497
L2	pu_exposure	Full covariates, survey weights	0.13	1.66	0.939	344	0.254407
L2	pu_exposure	Full covariates, no weights	-3.47	1.85	0.064	344	0.207463
L2	intervention_binary	Full covariates	21.62	12.71	0.092	344	0.204555
L2	pu_dexp13	Lagged school-level learning outcomes	-1.97	1.8	0.277	366	0.126481
L2	pu_dexp13	Lagged outcomes and covariates	-3.09	1.81	0.09	333	0.230121
L4	pu_exposure	Simple model with survey weights	1.41	1.34	0.295	376	0.006769
L4	pu_exposure	No survey weights but clustered SEs	6.78	0.64	0	3202	0.098513
L4	intervention_binary	Binary exposure variable	34.47	15.19	0.025	376	0.028946
L4	pu_exposure	Full covariates, survey weights	0.49	0.98	0.616	340	0.309592
L4	pu_exposure	Full covariates, no weights	-0.74	0.84	0.384	340	0.240669
L4	intervention_binary	Full covariates	23.01	11.09	0.041	340	0.249126
L4	pu_dexp13	Lagged school-level learning outcomes	0.96	1.17	0.414	364	0.168397
L4	pu_dexp13	Lagged outcomes and covariates	0.12	1.03	0.908	328	0.262732
N2	pu_exposure	Simple model with survey weights	1.23	3.57	0.732	369	0.001381
N2	pu_exposure	No survey weights but clustered SEs	11.62	1.47	0	2801	0.057978
N2	intervention_binary	Binary exposure variable	39.79	18.3	0.032	369	0.033379
N2	pu_exposure	Full covariates, Survey weights	2.76	2.42	0.258	334	0.240537
N2	pu_exposure	Full covariates, no weights	0.38	2.26	0.868	334	0.17945
N2	intervention_binary	Full covariates	44.46	15.58	0.005	334	0.215065
N2	pu_dexp13	Lagged school-level learning outcomes	-0.8	2.34	0.733	355	0.063586
N2	pu_dexp13	Lagged outcomes and covariates	1.01	2.19	0.647	323	0.201071
N4	pu_exposure	Simple model with survey weights	3.55	2.05	0.086	363	0.024344
N4	pu_exposure	No survey weights but clustered SEs	6.9	0.7	0	3177	0.085963
N4	intervention_binary	Binary exposure variable	64.82	20.55	0.002	363	0.054765

N4	pu_exposure	Full covariates, survey weights	2.27	1.83	0.22	331	0.235603
N4	pu_exposure	Full covariates, no weights	-0.76	1.61	0.638	331	0.148004
N4	intervention_binary	Full covariates	63.86	21.43	0.004	331	0.192514
N4	pu_dexp13	Lagged school-level learning outcomes	-1.25	1.95	0.523	350	0.012849
N4	pu_dexp13	Lagged outcomes and covariates	-0.56	2.05	0.786	318	0.146748