Oxford Policy Management

CAMBRIDGE EDUCATION

esspin
Education Sector Support Programme in Nigeria

UKaid
from the British people

ESSPIN Composite Survey 3

Jigawa State Report

Katharina Keck

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 and Shefali Rai provided helpful comments on an earlier draft of this report.

From the Education Sector Support Programme in Nigeria (ESSPIN), we acknowledge the contribution of Ron Tuck, who first devised the concept of the Composite Survey; and we also acknowledge the various leadership roles undertaken in relation Composite Survey 1 (CS1) and Composite Survey (CS2) by 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, Hezekiah Odeyale and Abubakar Mohammed Nashabaru provided valuable information and support. Allan Findlay and Agbenu Onazi provided information on enrolment and school numbers from the Annual School Censuses.

The Composite Surveys would not have been conducted without UK aid funding provided by the UK Department for International Development (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 Surveys. 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.

This survey is being carried out by Oxford Policy Management for the Education Sector Support Programme in Nigeria (ESSPIN), which is funded by the UK Department for International Development and managed by a consortium led by Cambridge Education. The survey technical lead is Stuart Cameron. The project manager is Shefali Rai. For further information please contact shefali.rai@opml.co.uk.

Oxford Policy Management Limited
Registered in England: 3122495

Level 3, Clarendon House
52 Cornmarket St
Oxford OX1 3HJ
United Kingdom

Tel +44 (0) 1865 207 300
Fax +44 (0) 1865 207 301
Email admin@opml.co.uk
Website www.opml.co.uk

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 Jigawa State from the first, second and third rounds of the ESSPIN 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 is to understand change in schools over time, and whether schools which receive intervention through ESSPIN are working better than those which do not. The main findings are as follows:

Head teacher effectiveness has not changed significantly over time in Jigawa. Around 17% of head teachers meet ESSPIN's overall standard for head teacher effectiveness in 2016. However, there has been a marked improvement in the aspects of effectiveness on which training has focused and which are most under head teachers' direct control. Overall, ESSPIN's intervention is not associated with better performance on the head teacher effectiveness standard.

School development planning in Jigawa has improved dramatically since 2012. Around 37% of schools meet the standard for effective school development planning in 2016, and schools which have had more years of ESSPIN intervention have better school development planning than those which have received fewer years of intervention.

Trends in **inclusiveness** are measured by aspects such as whether the head teacher has taken action on learners' attendance, and whether teachers engage boys and girls equally. In 2016, 14% of schools meet ESSPIN's standard for inclusiveness. This is a slight decline compared to 2012, but an improvement compared to 2014. ESSPIN intervention is not associated with greater inclusiveness in schools.

SBMCs in Jigawa have become much more functional since 2012, and are more likely to have conducted awareness raising about the value of education, and to have addressed exclusion and raised issues of children's exclusion with the Local Government Educational Authority (LGEA) or state government, than in 2012. In 2016, 68% of SBMCs meet the standard for a functioning SBMC, compared to 20% in 2012. SBMCs are no more inclusive or women and children within their committees than they were in 2012. ESSPIN intervention is associated with better functioning SBMCs, and with SBMCs being more inclusive of children.

Teachers trained through ESSPIN perform better on English and numeracy content knowledge tests than non-ESSPIN-trained teachers. Performance on these tests has worsened over time, but less so for ESSPIN-trained teachers. Teacher competence has improved slightly over time, but ESSPIN's intervention is not associated with more competent teachers. ESSPIN-trained teachers are more motivated and more engaged than non-ESSPIN-trained teachers.

Overall **school quality** has improved since 2012, according to our composite measure based on head teacher effectiveness, school development planning, SBMC functionality and teacher competence. In 2016, 26% of schools meet the standard for a good quality school, compared to only 2% in 2012. This translates into an estimated additional 430 good quality schools, which educate approximately 160,000 learners. Schools that have received more years of ESSPIN

intervention score higher on measures of school quality, and have improved faster over time, than schools that have received fewer years of ESSPIN intervention. However, no school in the sample was able to meet a stricter school quality standard, which includes half the teachers at a school passing a literacy and numeracy test.

Children's **learning outcomes** have improved over time in Grade 2 literacy and Grade 4 numeracy, with no change in Grade 4 literacy and Grade 2 numeracy. ESSPIN's intervention is associated with higher scores on the Grade 4 tests, even after controlling for pre-existing differences between schools. There is no effect of ESSPIN intervention on Grade 2 test scores.

Table 1, Table 2, Table 3 and Table 4 summarise the key findings.

Table 1: Jigawa: Change over time – Key indicators in 2012, 2014, 2016

	See page	2012 (CS1)	2014 (CS2)	2016 (CS3)	Change 2012–16	Change 2014–16
Effective head teacher (%)	15	16.9	3.8	16.6	-0.3	**+12.8**
School development planning (%)	17	1	6.2	36.6	**+35.6***	**+30.4***
Inclusive (%)	19	19.8	4.7	14.4	-5.4	+9.7
Functioning SBMC (%)	22	20.4	24.5	68.3	**+47.9***	**+43.8***
Competent teachers (%)	32	62.2	64.5	68.6	+6.5	+4.1
Competent teachers (new measure, %)	32		5.8	4.3	n/a	-1.5
Good quality school (%)	41	2.2	3.6	26.2	**+24.0***	+22.6
Good quality school (new measure, %)	41		0	0	n/a	0.0
Grade 2 literacy score	45	433.9	402.2	424.6	-9.3	**+22.4***
Grade 4 literacy score	45	411.5	405.4	400.1	-11.4	-5.3
Grade 2 numeracy score	45	445.1	421	413.6	-31.5	-7.4
Grade 4 numeracy score	45	405.7	385.8	429.2	+23.5	**+43.4***
Note. * indicates statistical significance (p < .05)						

Table 2: Jigawa: Key indicators in 2016, by ESSPIN Output Stream 3 intervention

	See page	Min. (1 year)	Med. (2–3 years)	Estimated effect of one year of full intervention
Effective head teacher (%)	15	18.4	14.1	+0.4
School development planning (%)	18	13.1	60.9	**+14.0***
Inclusive (%)	20	16.7	11.9	+2.0
Good quality school (%)	41	15.6	37.2	+10.3
Good quality school (new measure, %)	41	0	0	n/a
Grade 2 literacy score	46	428.5	422	+2.2
Grade 4 literacy score	46	394.7	404.5	**+17.9***
Grade 2 numeracy score	46	434.9	399.8	-1.2
Grade 4 numeracy score	46	418.3	438.1	+23.5
Note. * indicates statistical significance (p < .05).				

Table 3: Jigawa: Key indicators on SBMCs and inclusion, by ESSPIN Output Stream 4 intervention

	See page	No inter-vention	Post-CS1	Pre-CS1	Estimated effect of one year of intervention
School meets standard for functioning SBMC (%)	24	63.4	92.3	79.2	+12.4
SBMCs' work on inclusion	24				
Conducted awareness raising (%)		74.1	77.5	87.1	+3.6
Addressed exclusion (%)		75.2	79.2	85.2	+3.2
Took action for commonly excluded groups (%)		38.2	53.8	37.4	+4.1
Raised issues of children's exclusion (%)		34.7	46.8	33.5	+3.0

Note. * indicates statistical significance (p < .05). ESSPIN's Output Stream 4 intervention focuses on improving community participation in school improvement.

Table 4: Jigawa: Teacher competence and test performance, by ESSPIN training

	See page	Non-ESSPIN-trained	ESSPIN-trained	Difference
Competent teachers (%)	33	78.7	63.2	-15.5
Teacher competence standard (excl. curriculum knowledge, %)	33	95.4	92.2	-3.2
Competent teachers (strict measure) (%)	33	3.7	4.7	+1.0
Teachers' English scale	36	384.2	401.5	+17.3
Teachers' mathematics scale	36	419.4	431.9	+12.5

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
BEd	Bachelor of Education
CAPI	Computer-assisted personal interviews
CBOs	Community-based organisations
CS1	Composite Survey 1
CS2	Composite Survey 2
CS3	Composite Survey 3
DFID	Department for International Development (UK)
ESSPIN	Education Sector Support Programme in Nigeria
IRT	Item response theory
L2	Grade 2 literacy test
L4	Grade 4 literacy test
LGA	Local Government Authority
LGEA	Local Government Educational Authority
N2	Grade 2 numeracy test
N4	Grade 4 numeracy test
NCE	National Certificate of Education
OPM	Oxford Policy Management
PGDE	Post-Graduate Diploma in Education
PTR	Pupil–teacher ratio
SBMC	School-Based Management Committee
SD	Standard deviation
SDP	School Development Plan
SE	Standard error
SIP	School Improvement Programme
SMO	Social Mobilisation Officer
SSIT	State School Improvement Team
SSO	School Support Officer

SUBEB	State Universal Education Board
TDP	Teacher Development Programme

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 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 Jigawa 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' own 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, after which schools continue to receive school visits from government officers to help maintain and continue improving quality gains. At the same time, many of the same schools have been receiving interventions under the fourth output stream, facilitating community involvement and inclusion through SBMCs.

1.2 ESSPIN in Jigawa State

ESSPIN has been working in partnership with the government of Jigawa since 2009. A key distinctive feature of ESSPIN's involvement in Jigawa is that the mix of intervention activities has varied with the phase of roll-out. This means that, aside from some schools having received ESSPIN's intervention for more years than others, schools also differ on the mix of intervention they have received. This is likely to complicate the interpretation of results when comparing different intervention groups.

ESSPIN's support to schools in Jigawa has encompassed all three elements of the SIP – support to teachers, head teachers, and school improvement planning. Teachers have received training on basic teaching skills, classroom organisation, teaching aids and giving 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 School Support Officers (SSOs).

In addition to these interventions, since 2010/2011 schools in Jigawa have received support under ESSPIN's fourth output stream: 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, and child protection and participation. This has been complemented by follow-up mentoring and monitoring visits by SMOs.

The extent to which schools have received each of the intervention components has varied from one year to the next (see Annex B and Annex C). In terms of Output Stream 3 interventions (Annex B), 146 pilot schools started off receiving two years of 'full intervention' (leadership training, teacher training and school improvement visits), followed by one year of no intervention activity. In 2012/2013, a further 290 schools received a year of full intervention, and in 2013/2014 an additional 436 schools received a year of full intervention. In 2014/2015, ESSPIN's Output Stream 3 interventions were scaled up to all schools in the state, encompassing a further 1,065 schools. Each of these different phases has involved a different form of intervention delivery. For example, in 2014/2015, 436 schools received 20 days of leadership training, while the remaining schools in Jigawa received only between four and eight days. This means that the 'ESSPIN experience' has varied depending on the phase of roll-out.

Until recently, the delivery of Output Stream 4 interventions (Annex C) in Jigawa has been fragmented. Until 2013/14, no school had received the three different intervention components within the same year. In 2014/15, 400 schools received the full intervention package and a further 501 schools received SBMC training and mentoring visits, but no training on women's and children's participation. The Output Stream 4 interventions have not yet been fully scaled up in Jigawa, with the majority of schools (1,468) having received only five days of mentoring visits to date. Table 5 shows what proportion of schools in Jigawa received Output Stream 3 and Output Stream 4 intervention in any given year.

Table 5: Proportion of schools receiving ESSPIN Output Stream 3 and Output Stream 4 interventions

%	2009/10	2010/11	2011/12	2012/13	2013/14	2014/15	2015/16	Any year
Full package of Output Stream 3 interventions	8	8	0	15	37	100	55	100
Any Output Stream 4 intervention	0	10	10	10	52	100	52	100

Note: Proportions are calculated relative to the estimated total number of schools in 2015/16, and so these are not perfectly accurate for other years because the total number of schools changes slightly from year to year. Where

Annual School Census numbers are lower than ESSPIN's intervention tables, the information from ESSPIN is used on the assumption that there are some missing data in the Annual School Censuses.

The expansion of the programme to all schools in the state required a changed model for delivering training, with the training located closer to schools. During the pilot phase of ESSPIN (2009/10 and 2010/11), State School Improvement Teams (SSITs), trained directly by ESSPIN staff, were responsible for supporting and training head teachers and teachers directly. As the programme expanded (from 2012/13 on), the SSITs and ESSPIN-trained SSOs – local government-level staff – who in turn delivered the school-level interventions.

Stakeholders perceive that having Local Government Authority (LGA) officers deliver the SIP has brought support closer to the schools, and that a school's needs can therefore be addressed more directly. However, the change in model is likely to have affected the quality of implementation to some extent. Programme staff argue that locating training closer to the schools has longer-term benefits, but that in the shorter term the quality standards of the pilot programme might not be fully upheld as the new, much larger numbers of trainers, who typically have lower qualifications than those in the first wave, develop competencies.

1.3 Contextual factors and their implications for the SIP in Jigawa

This section describes key aspects of the backdrop against which ESSPIN's implementation in Jigawa has taken place in recent years. These contextual factors are relevant when interpreting the changes in school-level outputs and outcomes between CS1, 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:

1. contextual factors or changes that **affect all schools in the state**, whether positively or negatively. These factors may help explain trends across the state as a whole; and

2. contextual factors or changes that **particularly affect schools with more years of ESSPIN intervention**. These factors may interfere with the analysis of ESSPIN's impact.

Contextual factors that have affected all schools in the state

The most significant contextual change in the last couple of years has been the sharp increase in enrolment in Jigawa schools. This has not been accompanied by the recruitment of new teachers, and has therefore also led to sharp increases in pupil–teacher ratios (PTRs).

According to Annual School Census data, enrolment in the state increased by more than 16% between 2009 and 2013, but stabilised between 2013 and 2014 (Table 6). Between 2009/10 and 2014/15, 83,575 additional children were absorbed by the school system. Enrolment increases are likely to affect both the number and the composition of pupils, as learners from disadvantaged and conflict-affected backgrounds, who would previously have been excluded, are presumably among the new entrants to the system[1].

Between 2009 and 2014, average PTRs increased from 46 to 55 (see Annex A). Analyses at the school level suggest that, between 2013 and 2014, PTRs continued to increase in smaller schools, but decreased slightly in larger schools. Stakeholders report that there has been no recruitment of

[1] Little data are available on the profile of new learners entering schools. However, there have in the past been strong correlations between school attendance and socio-economic status (Mezger, 2016). This means that further increases in enrolment are likely to come disproportionately from households that are the poorest, and where the learner's parents have relatively low levels of education.

new teachers in Jigawa over the past two years. With teachers retiring and enrolment increasing, this has left schools short-staffed, and teachers facing more difficult teaching conditions.

In 2014/2015, enrolment in Jigawa fell short of gender equity, with only 43.2% of enrolled pupils being female (240,106 female pupils compared with 315,289 male pupils). Out of the six ESSPIN states, Jigawa has the lowest proportion of female pupils.

Table 6: Number of schools and enrolment in the 2009, 2013 and 2014 Annual School Censuses

	Enrolment	Number of schools	Enrolment change (%)
2009/10	471,820	1,868	
2013/14	550,813	1,997	+16.7
2014/15	555,395	2,012	+0.8
Overall			+17.7

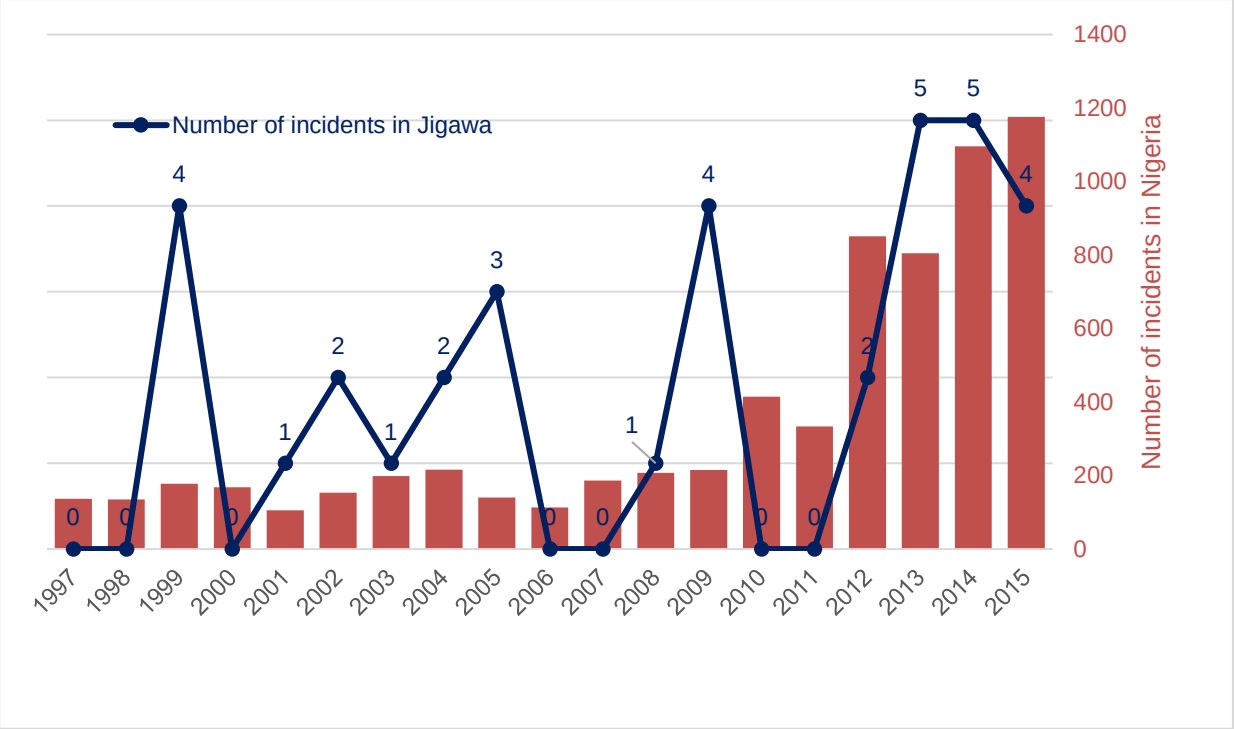
Note: Enrolment is for Primary Grades 1–6. Source: Annual School Census.

The Teacher Development Programme (TDP) also operates in Jigawa, and is administered through similar operational structures as ESSPIN. In particular, SSOs deliver the intervention activities to schools under both the TDP and ESSPIN's SIP. The implications of this for the delivery of the SIP are unclear. Stakeholders suggest that there have been difficulties at a managerial level, including training clashes and a lack of coordination of activities. However, stakeholders believe that the co-existence of the two programmes has had relatively little effect on the implementation of the SIP at the school level.

There also appear to have been some contextual changes that are supportive of school-level outcomes. There was a change in government after the 2015 elections, and according to stakeholders the new government has facilitated access to UBEC-IF funding. These changes may have occurred too recently to have had an impact that could be captured in this round of the survey. However, the implementation data do suggest that Output Stream 4 intervention delivery in particular has been more consistent and comprehensive over the last two years.

While insecurity remains a concern in Jigawa, the number of fatalities that are the result of political violence appears to be decreasing (see Figure 1 and Table 7). Stakeholders reported that there have been some community clashes and attacks by armed herdsmen, and that conflict surrounding the roll-out of the polio vaccine led to the temporary closure of a few schools in 2014/15. However, the impact of these events appears to not have been widespread and the implications for the school-level outcomes considered in the Composite Surveys are therefore likely to be fairly limited.

Figure 1: Incidents of political violence in Nigeria and Jigawa



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 7: Jigawa: Political violence: Incidents and fatalities, 2010–2015

Variable	2010	2011	2012	2013	2014	2015
Events	0	0	2	5	5	4
Fatalities	0	0	1	9	17	2

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

Over time, we would expect that schools in CS3 would have higher measures on our indicators than schools in CS1 and CS2. However, this section also highlights one key factor that may dampen the effect of ESSPIN's intervention to some extent: the substantial increase in enrolment and PTRs.

Contextual factors that disproportionately affect schools with more ESSPIN intervention

The schools that were selected for ESSPIN in the earlier phases differ significantly from those schools which were included only in the later expansion in 2014/15 (Annex A). The schools with more years of ESSPIN intervention are more likely to be urban, to have more qualified teachers and to have lower PTRs. These differences between the schools have not changed much over time: learner enrolment and PTRs have increased at similar rates in all schools in Jigawa. In 2014/15, schools with more years of ESSPIN intervention had an average PTR of 48, compared to an average PTR of 61 for schools that had received fewer years of intervention.

These differences make it difficult to isolate the effect of ESSPIN intervention. In particular, schools that have received more years of ESSPIN intervention up to now may also have been more likely to meet ESSPIN standards even before the start of the intervention due to their background

characteristics. Over time, better infrastructure and access to more resources may allow schools with more years of intervention to improve faster. These contextual factors should be kept in mind in the interpretation of the results presented in this report.

2 Methodology and analysis

2.1 Evaluation strategy

2.1.1 ESSPIN intervention groups

The original evaluation design for ESSPIN relied on maintaining a control group of schools with no intervention, which could be compared to those with a longer history of intervention (Phase 1: roll-out prior to the 2012/13 school year) and to those where intervention had started more recently (Phase 2: roll-out in 2012/13 or 2013/14). In practice, the roll-out followed a different implementation plan, with the result that by CS3 all schools in Jigawa had received some level of ESSPIN SIP activities.

While faster roll-out and greater reach may be a sign of success for the ESSPIN programme, it presents a difficulty for evaluation as there is now no longer a control group which has received no intervention. The nature, timing and intensity of ESSPIN intervention varies widely both between and within states (Annex B shows the number of days of leadership training, teacher training and school visits that schools have received under Output Stream 3; Annex C shows the number of days of SBMC training, training in women's and children's participation and mentoring visits received under Output Stream 4).

To simplify the analysis, we focus our analysis on the number of years of full Output Stream 3 intervention that schools have received. Full intervention means that the school has received some leadership training, some teacher training, and some school visits during the year, though the amount of each activity may vary. A further simplification groups the schools into minimum (zero to one year), medium (two to three years), and maximum (four to five years) intervention categories. We also assume that it takes a year for ESSPIN's interventions to have an impact. In line with this, we have not considered activities carried out during 2015/16 when grouping schools into different intervention categories as the results of these activities will not have emerged by the time of CS3.

For certain indicators, we alter the classification scheme slightly according to the purpose of our analysis. For example, when examining teacher competence within the CS3 survey, we consider two different groups: teachers who are in schools that have received ESSPIN's intervention but who have not themselves been trained by ESSPIN[2]; and teachers who have been trained by ESSPIN. When examining SBMC functionality and inclusive practices of SBMCs, we classify schools according to the amount of Output Stream 4 intervention received. Schools are classified as 'no intervention' (five or fewer days of Output Stream 4 intervention received), 'pre-CS1' (started receiving intervention in 2011/12 or prior to this), and post-CS1 (started receiving intervention in 2012/13 or after).

Learning outcomes – literacy and numeracy in Grades 2 and 4 – are analysed using item response theory (IRT), providing an overall scale of how well children have scored, as well as a grouping of children into levels corresponding to the level they are expected to reach by the end of each grade. The distributions of children's performance across different intervention groups, and across the different survey rounds, are compared. Scores in specific test sub-scales (in literacy these are

[2] Three to six selected teachers within each school attended workshops delivered by SSOs. In some states the same group of teachers continued to receive training year after year, while in other cases attempts were made to spread the training to teachers who had not yet received any. However, teachers in ESSPIN schools are also expected to receive more support through other channels, and particularly through professional development meetings organised by the head teacher (RTI International, 2014; and personal communications from ESSPIN). We distinguish the teachers who received direct training ('ESSPIN-trained') from those who were not themselves directly trained but who are in ESSPIN schools and so are expected to have received support from their head teachers and colleagues ('not ESSPIN-trained').

labelled *receptive*, *fluency in reading and writing*, and *productive*; in numeracy they are *calculation*, *everyday maths*, and *word problems*) are also analysed by intervention group and over time.

Teacher tests – literacy and numeracy – are analysed in a similar manner, also using IRT. As for learners, the analysis includes an overall scale of how well teachers have scored, groupings of teachers according to which grade level they have achieved, and analyses of specific sub-scales within the literacy and numeracy tests.

2.1.2 Types of analysis

The purpose of CS3 is to provide insights into the changes over time in the six states in which ESSPIN works, and to evaluate whether ESSPIN is having an effect in the specific schools in which its school improvement and community inclusion interventions have been applied. 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 same 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 at Grades 2 and 4. For each of these indicators we present in the following chapter two main types of analysis:

1. Change over time between CS1 and CS3, and between CS2 and CS3 in Jigawa. It is important to monitor change over time in how schools function and how much children are learning, both to inform programmes such as ESSPIN and for broader education policy-making. Trends over time in ESSPIN states are likely to reflect both the presence of the intervention and a number of other economic, social and political factors. If ESSPIN has been successful in this aim, then we would expect – other things being equal – that schools in CS3 will have higher output, outcome and impact measures than schools in CS1 and CS2. In practice, however, many other things may not be equal. Changes in enrolment, student profile, state financing, and political commitment may all affect these indicators at the same time. We present these changes over time and, where information is available, consider what may be driving changes, aside from ESSPIN intervention.

We use statistical significance tests (t-tests) to give an indication of whether a difference in results between our samples is likely to reflect a genuine difference in the overall populations. Given two 'populations' or groups of interest that we wish to compare – say, schools in Jigawa in 2012 and schools in Jigawa in 2016 – a common approach is to take a random sample from each group and compare the average performance in one sample to that in the other sample. However, there will be some random variation between the two samples that is due to the set of schools that happened to be sampled. This random variation could result in differences between the two samples even when the two populations are the same. Statistical tests tell us the probability that a difference between the two groups occurred by chance due to random variation in the samples, as opposed to being due to genuine differences in the two populations that the samples were drawn from. When we are looking at change over time, the t-test tells us the probability that a difference between our 2012 and 2016 sample is due to chance variation between the samples, as opposed to reflecting a genuine change over time in Jigawa's schools. A probability (sometimes known as the 'p-value') of 5% or less is often taken to be a good threshold for accepting that there is a genuine change, and we mark the result with an asterisk (*) when this is the case.

2. Differences between the different levels of intervention categories (minimum, medium and maximum) within the CS3 results. We hypothesise that schools that have received more years of full ESSPIN intervention will have higher output, outcome and impact measures than schools which have received fewer years of intervention.

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. We also show the averages for each intervention group. There are no maximum intervention schools in the Jigawa sample, so we compare results between the minimum (zero to one years) and medium (two to three years) intervention groups. In terms of Output Stream 4 intervention, we compare results between 'no intervention', 'post-CS1' and 'pre-CS1' intervention groups.

While it makes sense to compare the outcomes of schools with different levels of exposure to the intervention, there are likely to be spill-over effects between schools, which means that staff in the minimum intervention group might have already been exposed to ESSPIN ideas through informal communication, or deliberately by LGEA personnel. In particular, the School Improvement Model anticipates that ESSPIN-trained teachers should share their training with the rest of the school staff, with support from head teachers, and therefore spill-over effects to other teachers are likely.

Once again, any significant results cannot conclusively be attributed to the ESSPIN intervention. Schools, teachers and pupils that have received fewer years of ESSPIN intervention are likely to differ on background characteristics from those that have received more years of ESSPIN intervention, which is in turn likely to affect their performance. For example, as discussed earlier, schools which have received more years of ESSPIN intervention are typically more urban than schools which were included in the later roll-out (and which have therefore received fewer years of ESSPIN intervention). Controlling for this fully is a more difficult statistical exercise, so we will only attempt this for our outcome and impact measures: school quality and pupil learning outcomes.

For these indicators, we conduct additional analyses in order to understand what basis there might be for making causal attribution of ESSPIN's impact. This analysis is described in Sections 5 and 6.

2.2 Sampling, coverage and weights

In CS3, all the schools visited in CS2 were visited again, with the intention of collecting data that would enable us to analyse changes over time and differences between schools that have received different amounts of ESSPIN intervention.

The sample size for Jigawa has been 105 schools for all three rounds of the survey. In CS1 and CS2, 103 schools could be visited. For CS3, the same schools were visited as during CS2, along with two additional schools to bring the sample size back up to 105. Two schools from the CS2 sample were replaced, one because it no longer exists and one because it was found to be ineligible for the survey during CS2.[3]

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 8. Our sample includes only four schools that have received two years of ESSPIN intervention. Therefore, the results presented for the medium intervention group would not necessarily be an accurate reflection of the subset of schools that have received two years of ESSPIN intervention.

[3] This school was a special school for deaf children, with an adapted curriculum. It was therefore not comparable with other schools in Jigawa.

Table 8: Sample in CS1 and CS2 and population of schools, by Output Stream 3 intervention group

Intervention category	Years of intervention	CS1 sample	CS2 sample	CS3 sample	Population of school
Minimum	1	32	30	30	1,065
Medium	2	0	2	4	436
	3	71	71	71	436
Total		103	103	105	1,937

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.

Within each school, the survey team conducted interviews with the head teacher, the SBMC chairperson or deputy, teachers and pupils. As in CS2, we intended to sample six teachers per school, or all of the eligible teachers in schools with fewer than six teachers. For CS3, we attempted to find the same teachers as were interviewed during CS2, using their photographs and name information, and to interview them again to be able to assess changes over time, as well as rates of change in teacher competence, and test results, with more precision. In Jigawa, we were able to re-sample 53.9% of teachers sampled during CS2. In schools where we could not re-sample all teachers from CS2, we topped up the sample with eligible teachers selected randomly from the attendance register.

Overall, 100% of the targeted sample of teachers were interviewed and had their lessons observed (Table 9). The sample coverage for the teacher tests is slightly lower because some teachers did not give their consent to be tested, or were not able to stay for the teacher tests, which were held after lessons had finished

As in CS2, four pupils were sampled per school for each of the tests (Grade 2 literacy, Grade 2 numeracy, Grade 4 literacy and Grade 4 numeracy). Where possible, we sampled pupils who were taught by one of the sampled teachers. The sample coverage of pupils was 97.3%. 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: Jigawa: Survey instruments, respondents, sample size and coverage in CS3

	Respondents	Targeted sample size	Number of respondents covered	Sample coverage (% of targeted sample size)
Head teacher interview	Head teachers	105	105	100%
SBMC interview	SBMC chair person	105	105	100%
Teacher interview	Sampled teachers	462	462	100%
Literacy test	Sampled teachers	462	449	94%
Numeracy test	Sampled teachers	462	436	97%
Lesson obs	Sampled teachers	462	462	100%
L2	Sampled P2 pupils	420	408	97%
N2	Sampled P2 pupils	420	406	97%
L4	Sampled P4 pupils	420	411	98%
N4	Sampled P4 pupils	420	410	98%

Note. In this table and throughout this report, L2 refers to the Grade 2 literacy test, L4 to the Grade 4 literacy test, N2 to the Grade 2 numeracy test, and N4 to the Grade 4 numeracy test.

The Composite Survey samples were designed to ensure different intervention groups were covered in large enough numbers, and so some types of school are over-represented in the samples (as Table 8 shows). This means that we have to apply sample weights, which give a greater weight to the results in schools that are relatively under-represented in the survey. Applying the sample weights when we calculate averages from the survey gives us results that are representative for Jigawa as a whole. Sample weights were calculated for the CS1, CS2 and CS3 schools, teachers, and pupils.

2.3 Fieldwork and instruments

Fieldwork for CS3 was conducted using computer-assisted personal interviews (CAPI) during April–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 (described in detail in the CS3 Overall Technical Report). At the same time, we retained the questionnaire items required for comparability with previous rounds of the Composite Survey.

Data were collected on teacher competence, head teacher effectiveness, school development planning, inclusive practices in schools, SBMC functionality, teacher competence, teacher subject knowledge and learning outcomes of children in Grades 2 and 4 in English and mathematics. The following activities were carried out as part of the data collection:

- structured interviews with head teachers, SBMC chairpersons and teachers;
- teacher tests in English literacy and numeracy;
- lesson observations; and
- literacy and numeracy tests for pupils in Primary Grades 2 and 4.

The instruments were pre-tested over 2 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.

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.

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 Jigawa 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 a school development plan (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, around 17% of head teachers in our sample met ESSPIN's standard for an effective head teacher. This does not represent a significant change compared to 2012 or 2014.
- In 2016, head teachers were more likely to have carried out lesson observations and to have held professional development meetings, but fewer schools were adhering to an appropriate lesson length.
- Head teachers in schools that had received more years of ESSPIN intervention by 2016 were no more effective than head teachers in schools that had received fewer years of ESSPIN intervention.

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) a 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).

There has been no significant change in the share of head teachers that meet ESSPIN's standard for effectiveness between 2012 and 2016 (Table 10). Over this period, some indicators have improved. Notably, the share of head teachers who conduct regular lesson observations and who hold regular professional development meetings with teachers have both increased dramatically. It is worth noting that these are the areas which have been the main focus of ESSPIN training. However, this has been offset by a sharp decline in the proportion of schools that adhere to an appropriate morning break, which worsened between 2012 and 2014, and continued to worsen between 2014 and 2016.

Fewer schools also conformed to a 35-minute lesson length in 2016 than in 2012 (measured as a lesson length of between 30 and 40 minutes). A length of 35 minutes was formerly considered the standard lesson length across the six states. However, schools have been encouraged to adopt 60-minute lessons, in line with ESSPIN lesson plans, which are intended to be taught over one hour. Longer lessons should therefore arguably be discounted as an indicator of school quality, as they may reflect a shift towards one-hour lessons in literacy and numeracy. We have therefore calculated a new indicator, defined as the proportion of schools in which at least half of the observed lessons are at least 30 minutes in length. On this indicator, schools also declined in 2016 compared to 2014, suggesting that lessons are in fact becoming shorter rather than longer. This may partly reflect the effect of observation on teachers. For example, they may be teaching components of lesson plans discussed during training, in a bid to impress the observer, but they may be unable to work these ideas into a full lesson. Even if this is the case, the short lesson times suggest that teachers have difficulty in planning lesson activities that fill a set duration.

Schools have worsened since 2012 in terms of our indicator of a clear opening time – the extent to which learners and teachers agree on what the school's opening time is. This appeared to be a result of confusion among learners in particular: on average, only 24% of learners in each school could agree on its opening time. The reasons for the worsening over time are not clear. However, it can be questioned whether this is a good indicator of school management. Field observations suggested that children were confused over 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 low, at 17%. In 2016, in only 40% of schools were at least half of the teachers and learners in class on time in the morning, suggesting that teachers' presence in the classroom remains a serious

concern. On other indicators, head teachers in Jigawa are performing well: 84% had carried out at least two lesson observations in the past two weeks, and 76% had taken at least two actions to promote teacher attendance. It appears that head teachers are performing well on factors which are in their direct control, but they have been less successful at promoting positive behaviour changes on the part of teachers in their schools.

Table 10: Jigawa: Head teacher effectiveness in CS1, CS2 and CS3

	2012 (CS1)	2014 (CS2)	2016 (CS3)	Change 2012-16	Change 2014-16
(1) Lesson observations (%)	6.8	10.1	83.4	**+76.7***	**+73.3***
– No. lesson observations in past two weeks	0.7	1	3	**+2.4***	**+2.0***
(2) Professional development meetings (%)	6.7	20.3	59.4	**+52.7***	**+39.0***
– No. professional development meetings last term	0.8	1.1	1.8	**+0.9***	**+0.7***
(3) Action on teacher attendance (%)	80	39.9	75.5	-4.5	**+35.6***
(4) Clear opening time (%)	41.5	19.4	4.8	**-36.8***	**-14.6***
– Learners who agree on opening time (%)		53.1	24.4	n/a	**-28.6***
– Teachers who agree on opening time (%)		72.9	74.7	n/a	+1.8
(5) In class on time in morning (%)	52.9	32.1	39.8	-13.0	+7.7
(6) Appropriate morning break (%)	89.8	73.4	52.6	**-37.1***	-20.7
(7) 35-minute lesson length (%)	61.5	60.4	23.8	**-37.6***	**-36.6***
– Lesson length 30 min or longer (%)		61	31.1	n/a	**-29.9***
Number of criteria fulfilled (out of seven)	3.3	2.6	3.4	+0.1	**+0.8***
Effective head teacher (five out of seven criteria met) (%)	16.9	3.8	16.6	-0.3	+12.8
Note. * indicates statistical significance (p < .05)					

Are the schools that have had more years of intervention under ESSPIN's Output Stream 3 doing better than those which have received the intervention only recently? We present averages across the two intervention groups: those that have only had one year of full intervention up to 2014–15 (minimum), and those that have had two to three years (medium) (Table 11). We also estimate the effect of one full year of intervention.

Overall, there is no difference in the proportion of effective head teachers between the minimum and medium intervention groups. Across all indicators, the estimated effect of ESSPIN intervention is not statistically significant. Schools with more years of ESSPIN intervention are somewhat more likely to have a clear opening time and head teachers are more likely to hold professional development meetings, but they are less likely to have conducted lesson observations.

Table 11: Jigawa: Head teacher effectiveness in 2016, by ESSPIN intervention

	Min. (one year)	Med (two to three years)	Estimated effect of one year of full intervention
(1) Lesson observations (%)	86.1	80.6	-6.1
– No. lesson observations in past two weeks	3.7	2.4	**-0.6***
(2) Professional development meetings (%)	46.2	73.1	+8.5

– No. professional development meetings last term	1.3	2.2	+0.3
(3) Action on teacher attendance (%)	68	83.3	+3.6
– School has a teacher attendance book (%)	100	100	n/a
(4) Clear opening time (%)	0.3	10.8	+9.4
– Learners who agree on opening time (%)	23	26.4	+3.6
– Teachers who agree on opening time (%)	76.5	72.8	-2.2
(5) In class on time in morning (%)	32.4	47.6	+3.2
– Classes where learners present on time (%)	97.4	95.2	+0.2
– Classes where teachers present on time (%)	35	44.9	+2.3
(6) Appropriate morning break (%)	65.5	39.3	-3.1
(7) 35-minute lesson length (%)	34.3	12.9	-7.3
– Lesson length 30 min or longer (%)	44.7	17	-9.0
Number of criteria fulfilled (out of seven)	3.3	3.4	+0.1
Effective head teacher (five out of seven criteria met) (%)	18.4	14.1	+0.4
Additional indicators			
(A1) In class on time after break (%)	41.2	82.3	**+16.2***
-- Classes where learners present on time (%)	95.5	98.8	+1.3
-- Classes where teachers present on time (%)	43.2	62.7	**+7.4***
(A3) Teacher absenteeism (%)	18.5	28.5	-4.3
Note. * indicates statistical significance (p < .05)			

3.2 School development planning

Box 3: School development planning: Key findings

- In 2016, schools had made substantial progress in school development planning, with large and statistically significant increases for four of five measured indicators.
- In 2016, we classify 37% of schools as effective at school development planning, compared to only 1% of schools in 2012. Despite the considerable improvement, the proportion of effective schools remains relatively low.
- Schools that have received more years of ESSPIN intervention are better at school development planning than schools that have received fewer years of ESSPIN intervention.

ESSPIN's leadership training encourages and supports head teachers to review their school's performance and put together a 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 that 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).

In 2016, schools had made substantial progress in their development planning, with large and statistically significant improvements for most indicators compared to 2012 and 2014 (Table 12). Schools' effectiveness in development planning improved between 2012 and 2014, and has continued to improve substantially between 2014 and 2016. Overall, 37% of schools met the standard for school development planning in 2016. While still fairly low, this is a massive improvement on 2012, when only 1% of schools met the standard.

As ESSPIN has been rolled out to all schools in Jigawa, school development planning appears to have become a much more widespread practice. Over 84% of schools had conducted a self-evaluation, and 82% had a school development plan available, compared to less than 10% of schools in 2012. The proportion of schools that have carried out at least four activities in their SDP has also improved over time, but the proportion of schools meeting this indicator remains low, at 22%. There is a need to engage with this finding, as the implementation of activities listed in the SDP is key if improvements in school development planning are to be translated into better school-level outcomes. It is possible, however, that some schools prefer to focus on a smaller number of high-impact measures.

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

	2012 (CS1)	2014 (CS2)	2016 (CS3)	Change 2012–16	Change 2014–16
(1) Written evidence of school self-evaluation process (%)	9.6	23.1	84.5	+74.8*	+61.4*
(2) SDP available (%)	9.9	21.7	82.3	+72.3*	+60.6*
(3) SDP contains three or more activities to strengthen teaching and learning (%)	2.3	6.2	51.7	+49.5*	+45.6*
-- No. activities in SDP to strengthen teaching and learning	0.1	0.6	2.4	+2.3*	+1.8*
(4) Evidence that four or more activities stated in SDP carried out (%)	1	6.4	21.8	+20.8	+15.3
-- No. activities in SDP carried out	0.1	0.5	2.2	+2.1*	+1.7*
(5) Cashbook up-to-date (%)	5.9	11.9	49.4	+43.5*	+37.5*
-- School has a cashbook (%)	11.7	26.3	85.9	+74.2*	+59.6*
Number of SDP criteria fulfilled (out of five)	0.3	0.7	2.9	+2.6*	+2.2*
School meets effective school development planning standard (four out of five criteria met) (%)	1	6.2	36.6	+35.6*	+30.4*

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

Schools with more years of ESSPIN intervention are more effective at school development planning than schools with fewer years of intervention. Almost 61% of schools in the medium

intervention group meet the standard for school development planning, compared to only 13% in the minimum intervention group. We estimate that one year of ESSPIN intervention is associated with an additional 0.3 criteria met.

Schools with more years of ESSPIN intervention are more likely to have carried out at least four activities stated in their SDP. They also perform better on the other indicators of school development planning, but the differences are not statistically significant. As noted above, this is not an estimate of ESSPIN's causal impact but lends support to the hypothesis that longer exposure to ESSPIN intervention has a positive effect on school development planning.

Table 13: Jigawa: SDP effectiveness in 2016, by ESSPIN intervention

	Min.	Med.	Estimated effect of one year of full intervention
(1) Written evidence of school self-evaluation process (%)	80.2	88.8	+0.4
(2) SDP available (%)	77	87.7	+1.2
(3) SDP contains three or more activities to strengthen teaching and learning (%)	34.8	69.4	+7.2
-- No. activities in SDP to strengthen teaching and learning	2.1	2.8	+0.1
(4) Evidence that four or more activities stated in SDP carried out (%)	9.1	34.9	**+11.2***
-- No. activities in SDP carried out	1.5	3	**+0.7***
(5) Cashbook up-to-date (%)	31.8	67.7	+11.1
-- School has a cashbook (%)	81.3	90.6	+3.2
Number of SDP criteria fulfilled (out of five)	2.3	3.5	+0.3
School meets effective school development planning standard (%)	13.1	60.9	**+14.0***
Note. * indicates statistical significance (p < .05)			

3.3 School inclusiveness

Box 5: School inclusiveness: Key findings

- Schools in Jigawa are slightly less inclusive than they were in 2012, but they are more inclusive compared to 2014.
- In 2016, only 14% of schools fully met ESSPIN's standard for an inclusive school.
- ESSPIN's intervention is not associated with schools being more inclusive.

The criteria relating to 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).

In 2016, the proportion of schools meeting ESSPIN's standard for an inclusive school has declined slightly compared to 2012, but has improved compared to 2014 (Table 14). Compared to 2012, schools have taken fewer actions on learner attendance, although they have taken significantly more compared to 2014. Similarly, teachers are less likely to use at least two assessment methods compared to 2012, but more likely to do so compared to 2014.

One indicator which has improved significantly since 2012 is the proportion of schools that include at least two activities on improving access in their SDP. This may reflect general improvements in school development planning, as noted in the previous section.

In the CS2 report, we were not able to identify a reason for the decline in inclusiveness in Jigawa's schools. However, as ESSPIN has been scaled up across Jigawa, schools appear to be becoming more inclusive. Despite this, the proportion of inclusive schools remains low, and only 14% of schools are able to fully meet ESSPIN's inclusiveness standard.

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

	2012 (CS1)	2014 (CS2)	2016 (CS3)	Change 2012–16	Change 2014–16
(1) Three or more actions on learner attendance (%)	66.2	11.8	58.3	-7.9	**+46.5***
– Number of actions on learner attendance	3.2	1.7	2.7	**-0.5***	**+0.9***
(2) Two or more activities in SDP on access (%)	1.2	13.5	37.4	**+36.2***	+23.9
– Number of activities on access	0.1	0.4	1.2	**+1.1***	**+0.8***
(3) >50% of teachers use two or more assessment methods (%)	59.5	27.8	38.4	**-21.1***	+10.6
(4) >50% of teachers spatially inclusive and >50% are gender inclusive (%)	31.6	15.8	23.4	-8.2	+7.6
Number of inclusiveness criteria fulfilled (out of four)	1.6	0.7	1.6	-0.0	**+0.9***
Inclusiveness score (%)	50.1	38.1	40.5	**-9.6***	+2.4
School fully met standard (three to four criteria) (%)	19.8	4.7	14.4	-5.4	+9.7
School partially met standard (two to four criteria) (%)	49.3	23.3	45.2	-4.1	+21.9

Note. * indicates statistical significance (p < .05). The inclusiveness score is a total ranging from 0 to 100 and is calculated as follows: $\mathbf{20}(\frac{s_1}{7} + \mathbf{min}\left(\mathbf{1}, \frac{s_2}{5}\right) + \frac{s_3}{3} + \frac{s_4}{6} + \boldsymbol{s_5})$, where s_1 is the number of actions to improve attendance; s_2 is the number of activities in the SDP to improve access for disadvantaged children; s_3 is the average number of assessment methods used by sampled teachers; s_4 is the average number of classroom zones participating in the lesson during lesson observations, and s_5 is the gender equity score (see below).

On most of the inclusiveness indicators, schools which have received more years of ESSPIN intervention perform slightly better than schools which have received fewer years of intervention, but the differences are small and non-significant. Schools that have received more years of ESSPIN intervention do appear to be more likely to have taken at least three actions on learner attendance.

Table 15: Jigawa: School inclusiveness in 2016, by ESSPIN intervention

	Min.	Med.	Estimated effect of one year of full intervention
(1) Three or more actions on learner attendance (%)	43.9	73.3	+9.9
– Number of actions on learner attendance	2.4	2.9	+0.2
(2) Two or more activities in SDP on access (%)	35.8	39.1	-2.2
– Number of activities on access	1.1	1.2	0.0
(3) >50% of teachers use two or more assessment methods (%)	48.5	27.9	-0.5
(4) >50% of teachers spatially inclusive and >50% are gender inclusive (%)	29	17.6	+0.5
Number of inclusiveness criteria fulfilled (out of four)	1.6	1.6	+0.1
Weighted sum inclusiveness score	38.9	42.1	+1.8
School fully met standard (three to four criteria) (%)	16.7	11.9	+2.0
School partially met standard (two to four criteria) (%)	45.4	45	+2.1
Additional indicators			
Enrolment increased since last year (%)	54.3	53.1	+6.3
Change in enrolment since last year	0	0.1	+0.1
Note. * indicates statistical significance (p < .05)			

3.4 SBMCs

Box 7: SBMCs: Key findings

- In 2016, SBMCs in Jigawa were much more functional than SBMCs in 2012 and 2014. About 68% of schools met ESSPIN's standard of SBMC functionality in 2016, compared to only 20% in 2012.
- ESSPIN's Output Stream 4 intervention is associated with better functioning SBMCs. SBMCs meet an estimated additional 0.6 criteria per year of ESSPIN's Output Stream 4 intervention they receive.
- Participation of women and children within SBMCs has not changed significantly since 2012, and the proportion of schools meeting each standard remains low (18% for women's inclusiveness, 4% for children's inclusiveness in 2016).
- SBMCs in schools that have had greater exposure to ESSPIN Output Stream 4 intervention are more inclusive of children, but only marginally more inclusive of women than SBMCs in schools that have had less exposure.

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 the 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 8). 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). Thus, they reflect the ability of the SBMC to keep good records of their activities, as well as actually undertaking the activities themselves.

Box 8: 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 community-based organisations (CBOs), traditional or religious institutions, or other SBMCs (written or physical evidence);
5) SBMC interacted with local government education authorities 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 from the start of the current school year (written evidence).

In 2016, SBMCs in Jigawa were much more functional compared to 2012 and 2014. About 68% of schools met the standard for a functioning SBMC, meeting an average of 5.1 out of nine criteria. This compares to 20% of schools meeting the standard in 2012, with an average of 2.8 of nine criteria met.

There have been large and significant increases since 2012 and/or 2014 in the proportion of SBMCs that met at least twice in the current school year and that have contributed resources for the school. Almost all schools (96%) had networked with other organisations. Many more had women's and children's committees.

SBMCs appear increasingly to be taking actions to improve access for the inclusion of all children (see Box 9 below). In 2016, a much higher proportion of SBMCs have conducted awareness-raising and have addressed exclusion compared to 2012. The proportion of SBMCs that took action for commonly excluded groups, such as children with disabilities, girls or nomadic children, and that raised issues of exclusion with the LGEA also increased over time. In 2016, 77% of SBMCs had addressed exclusion in some form, while 39% had specifically taken action for commonly excluded groups, and 35% had raised issues of exclusion with the LGEA.

[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: Jigawa: SBMC functionality in CS1, CS2 and CS3

	2012 (CS1)	2014 (CS2)	2016 (CS3)	Change 2012–16	Change 2014–16
(1) Two or more meetings this school year (%)	24.7	20.6	57.2	**+32.5***	**+36.6***
(2) Conducted awareness raising (%)	51.8	40.8	76.1	**+24.4***	**+35.3***
(3) Addressed exclusion (%)	31.9	49.6	76.8	**+44.9***	+27.3
(4) Networked with CBOs/institutions/other SBMCs (%)	27.1	61.9	95.7	**+68.5***	**+33.7***
(5) Interacted with LGEA (%)	22.3	16.4	38.5	+16.2	+22.1
(6) Has a women's committee (%)	20.6	15.3	46.3	+25.7	**+31.0***
(7) Has a children's committee (%)	15	14	36.7	+21.8	+22.7
(8) Contributed resources for school (%)	39.4	41.7	78.5	**+39.1***	**+36.9***
(9) Chair visited school three or more times (%)	41.6	13.2	17.1	-24.5	+3.9
Number of SBMC functionality criteria met (out of nine)	2.8	2.9	5.1	**+2.3***	**+2.2***
School meets standard for functioning SBMC (%)	20.4	24.5	68.3	**+47.9***	**+43.8***
Additional indicators: inclusion and drop-out					
(A1) Took action for commonly excluded groups (%)	22.8	21.2	39	+16.1	+17.8
(A2) Raised issues of children's exclusion (%)	3.9	20.3	35.2	**+31.3***	+14.9
Additional indicators: organising and mobilising resources					
(A7) School has an SBMC (%)	91.2	100	100	**+8.8***	+0.0
(A8) Cashbook available (%)	15.8	35.3	87	**+71.2***	**+51.7***
(A9) Requested support from LGEA or State Universal Basic Education Board (SUBEB) (%)			65	n/a	n/a
(A10) Raised cash to support school improvement (%)	11.6	18.7	34.3	+22.6	+15.6
(A11) Mobilised non-cash resources (%)	31.1	35.9	61.7	**+30.7***	+25.8
(A12) Involved in making SDP (%)		12.5	74.2	n/a	**+61.7***
Note. * indicates statistical significance (p < .05)					

Box 9: 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

ESSPIN's Output Stream 4 intervention is associated with better functioning SBMCs. On the individual criteria, results are in the expected direction for all nine indicators, but are mostly non-significant. This is in part because the 'post-CS1' group is generally performing better than the 'pre-CS1' group. While the 'pre-CS1' group has had Output Stream 4 intervention for longer (the first seven days of training were delivered in 2010/2011), overall the 'pre-CS1' and 'post-CS1' schools have had very similar amounts of intervention. In particular, both groups have received full intervention in the last two years, and have also received the same level of Output Stream 3

intervention. The more relevant comparison would therefore be between the 'no intervention' group and the other two groups.

SBMCs in schools with greater exposure to ESSPIN intervention are significantly more likely to have women's and children's committees.

Overall, 79% of SBMCs in schools that received Output Stream 4 intervention prior to CS1 meet ESSPIN's standard for a functioning SBMC, while 92% of SBMCs in the 'post-CS1' group meet the standard. This compares to 63% of SBMCs in the group that has received only five days of intervention so far ('no intervention'). We estimate that a year of full Output Stream 4 intervention is associated with 0.6 additional criteria being met.

SBMCs in schools with more Output Stream 4 intervention are slightly better at promoting the inclusion of all children, but results for these inclusiveness measures are statistically non-significant.

Table 17: Jigawa: SBMC functionality in 2016, by ESSPIN intervention

	No inter-vention	Post -CS1	Pre-CS1	Estimated effect of one year of full intervention
(1) Two or more meetings this school year (%)	54.4	78.6	44	+1.4
(2) Conducted awareness raising (%)	74.1	87.1	77.5	+3.6
(3) Addressed exclusion (%)	75.2	85.2	79.2	+3.2
(4) Networked with CBOs/institutions/other SBMCs (%)	94.9	100	95.4	+1.5
(5) Interacted with LGEA (%)	32	68.9	55.7	+12.5
(6) Has a women's committee (%)	36.9	88	76.3	**+24.6***
(7) Has a children's committee (%)	28.9	76.9	48.2	**+13.0***
(8) Contributed resources for school (%)	76.7	87.2	82.8	+4.1
(9) Chair visited school three or more times (%)	16.8	17	21.5	+1.3
Number of SBMC functionality criteria met (out of nine)	4.8	6.8	5.7	**+0.6***
School meets standard for functioning SBMC (%)	63.4	92.3	79.2	+12.4
Additional indicators: inclusion and drop-out				
(A1) Took action for commonly excluded groups (%)	38.2	37.4	53.8	+4.1
(A2) Raised issues of children's exclusion (%)	34.7	33.5	46.8	+3.0
(A3) Raised cash to support vulnerable children (%)	53.3	55.5	52.5	+0.2
(A4) Monitored drop-out or non-attendance (%)	53.1	66.4	74.4	+8.9
(A5) Communicated with school or community about drop-out (%)	96.7	100	94.5	0.0
(A6) No. actions taken to address non-attendance	2.1	2.6	2.3	+0.2
(A7) School has an SBMC (%)	100	100	100	n/a
(A8) Cashbook available (%)	84	100	97.1	+18.4
(A9) Requested support from LGEA or SUBEB (%)	61.4	80.8	76	+8.6
(A10) Raised cash to support school improvement (%)	29.2	55.7	53.2	+10.3
(A11) Mobilised non-cash resources (%)	57.7	78.8	75.6	+10.0
(A12) Involved in making SDP (%)	72.8	81.2	77.1	+2.9
Note. * indicates statistical significance (p < .05)				

3.4.1 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 10).

Box 10: 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]

SBMCs' inclusiveness of women has been improving marginally over time, but findings on most of the indicators are not statistically significant. The proportion of women's committees that have met and the attendance of women at SBMC meetings remain low and have not increased significantly over time. However, when female members do attend meetings, they are significantly more likely to raise issues: in 2016, female members had raised an issue in 52% of SBMCs, compared to only 22% of SBMCs in 2012. These issues are also somewhat more likely to have been addressed. This is consistent with qualitative research conducted for ESSPIN. While studies in 2009–2011 found that women's participation in SBMCs was very limited (Poulsen, 2009), by 2014, women felt that their suggestions were being heard in SBMC meetings (Little and Pinnock, 2014).

SBMCs' inclusiveness of children has not changed over time. Children's committees appear more likely to have met, and actions raised by children are more likely to have resulted in action in 2016 compared to 2012. The proportion of SBMCs where a child has raised an issue, however, remains low (13%), and has deteriorated marginally over time.

Levels of inclusiveness were low in 2012 and continue to be low in 2016. On average, schools meet less than one of the children's inclusiveness criteria. Only 18% of schools meet the standard for women's inclusiveness, and 4% of schools meet the standard for children's inclusiveness. The results point towards severe shortfalls remaining in regard to these indicators.

[5] This criterion has been slightly altered since CS1, where it also required that the women's committee 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.

Table 18: Jigawa: SBMCs' inclusion of women and children in CS1, CS2 and CS3

	2012 (CS1)	2014 (CS2)	2016 (CS3)	Change 2012–16	Change 2014–16
Participation of women in SBMC					
(1) At least one woman attended two or more meetings (%)	17.6	7.3	11	-6.6	+3.7
(2) Female member raised an issue (%)	22.3	29	51.5	**+29.3***	**+22.5***
(3) Issue raised by female member led to action (%)	19.3	11.7	29.8	+10.5	+18.1
(4) Women's committee met (%)	5.9	31.6	28.9	+23.0	-2.6
No. criteria met (out of four)	0.6	0.6	1	+0.4	+0.4
Meets standard (three out of four criteria) (%)	7.9	11.5	18.4	+10.4	+6.9
Participation of children in SBMC					
(1) At least one child attended two or more meetings (%)	12.5	4.3	11.2	-1.3	+6.9
(2) A child raised an issue (%)	16	19.4	13.3	-2.6	-6.1
(3) Issue raised by child led to action (%)	12.3	7.9	28	+15.6	+20.1
(4) Children's committee met (%)	1.9	4.5	23.7	+21.8	+19.1
No. criteria met (out of four)	0.4	0.4	0.7	+0.3	+0.3
Meets standard (three out of four criteria) (%)	3.5	4.1	3.8	+0.3	-0.3

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

More years of ESSPIN Output Stream 4 intervention are associated with greater inclusiveness of children's concerns within SBMCs. About 15% of SBMCs in schools in the 'pre-CS1' group are able to meet the children's inclusiveness standard, while no school is able to meet it in the 'no intervention' group.

There is a small but non-significant effect of ESSPIN intervention on meeting the women's inclusiveness standard. In SBMCs in schools with greater exposure to ESSPIN Output Stream 4 intervention, women's committees are more likely to have met and women are more likely to have attended SBMC meetings.

Table 19: Jigawa: SBMCs' inclusion of women and children in 2016, by ESSPIN intervention

	No intervention	Post-CS1	Pre-CS1	Estimated effect of one year of full intervention
Participation of women in SBMC				
(1) At least one woman attended two or more meetings (%)	7.2	29.9	17.9	**+4.8***
(2) Female member raised an issue (%)	54.2	47.2	42.1	-4.5
(3) Issue raised by female member led to action (%)	27.4	40	37.9	+4.7
(4) Women's committee met (%)	22.1	59.9	48.4	**+11.8***
No. criteria met (out of four)	0.8	1.7	1.4	+0.3
Meets standard (three out of four criteria) (%)	14.7	36.9	24.6	+5.4
Participation of children in SBMC				
(1) At least one child attended two or more meetings (%)	8.1	27.1	16.6	**+4.1***
(2) A child raised an issue (%)	9.6	17.2	53.8	**+8.2***
(3) Issue raised by child led to action (%)	28.2	24.4	33.1	+0.7
(4) Children's committee met (%)	17.1	61.3	28.4	**+9.1***
No. criteria met (out of four)	0.6	1.2	1	+0.2
Meets standard (three out of four criteria) (%)	0	22.1	15	**+3.5***
Note. * indicates statistical significance (p < .05)				

3.5 Summary: school management and head teachers

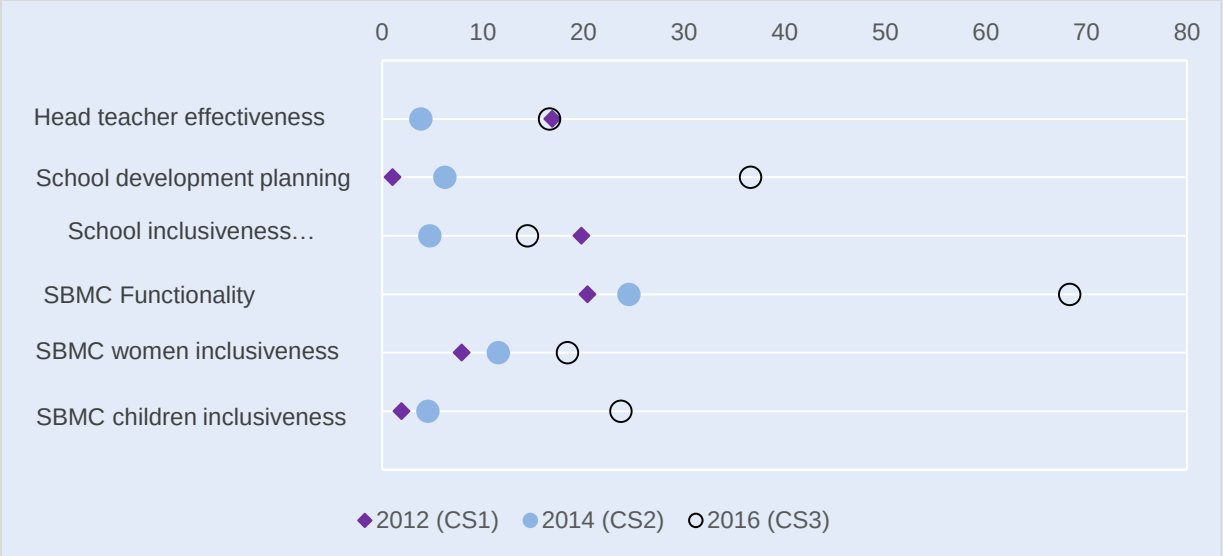
Box 11 shows the average progress in the school management standards in Jigawa primary schools in CS1, CS2 and CS3. The findings are mixed. Between 2012 and 2016, schools have improved substantially on their effectiveness in regard to school development planning, and SBMCs have become much more functional. About 37% of schools meet the standard on school development planning. Most schools are now conducting self-evaluations and preparing SDPs, but the proportion of schools who are able to implement at least four of the activities suggested in the SDP is lower. About 68% meet the standard for a functioning SBMC, and most SBMCs are conducting awareness raising, networking with other organisations and contributing resources to the school.

On the other hand, head teacher effectiveness, school inclusiveness and participation of women and children in SBMCs have not changed compared to 2012. Head teachers are, however, more likely to conduct lesson observations and hold professional development meetings than they were in 2012, and these were two key aspects that head teacher training focused on. Since 2012, all schools in our sample have received Output Stream 3 intervention, and most have received Output Stream 4 intervention, so a lack of progress on some of the school management indicators is worth investigating further. Contextual factors may also contribute to limited progress. For example, head teachers may be finding it difficult to improve some of their school management practices in light of increasing enrolment and limited resources. Out of the six ESSPIN states, in this round of the survey, Jigawa was the state in which most head teachers also participated in the teacher interview. It may be that head teachers have to invest more time in teaching themselves than focusing on school management. In particular, we find that practices in the head teacher's direct

control (such as developing an SDP or conducting lesson observations) have improved over time, but schools have been less successful at promoting positive behaviour changes for teachers in their schools (e.g. there are no changes in teacher presence in the classroom or the use of multiple assessment methods).

Box 11: Jigawa: School planning and SBMC functionality are improving over time

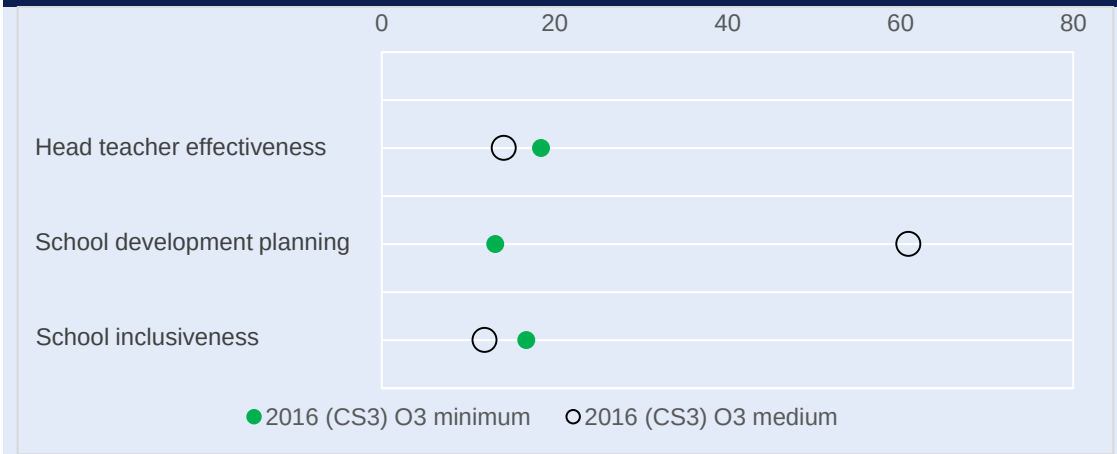
Comparison of school management indicators between CS1 (2012), CS2 (2014) and CS3 (2016)



Schools which have had more years of ESSPIN Output Steam 3 intervention perform better than schools with fewer years of intervention on school development planning, but not on head teacher effectiveness or inclusion (Box 12). Schools with more ESSPIN Output Steam 4 intervention perform better on SBMC functionality and children's inclusiveness, and somewhat better on women's inclusiveness. As shown in Box 13, these differences between the intervention groups are quite small. The 'post-CS1' group performs slightly better than the 'pre-CS1' group, but these schools have had a similar amount of Output Steam 4 intervention, although at different times.

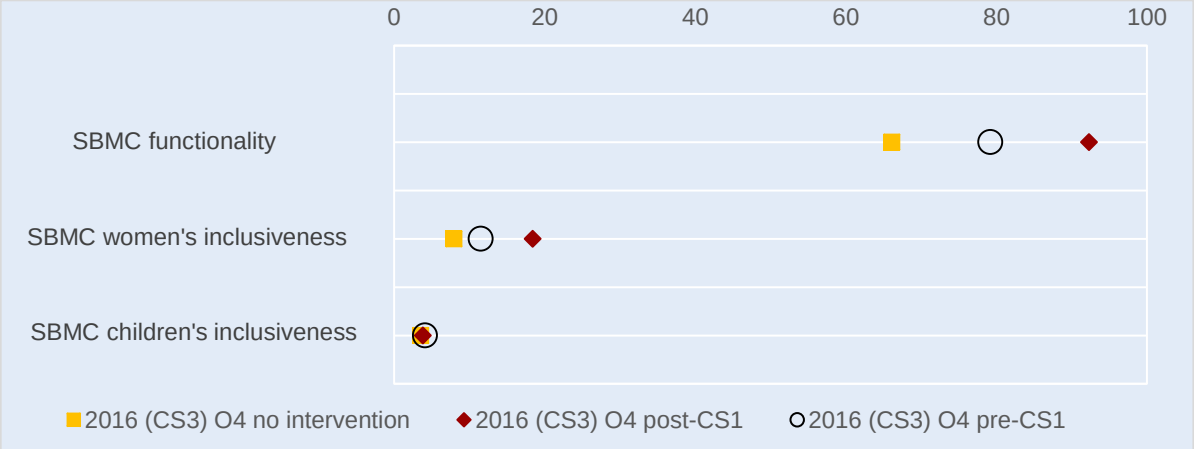
Box 12: Jigawa: Schools with more years of Output Steam 3 intervention are better at school development planning

Comparison of school management in minimum and medium intervention schools in 2016



Box 13: Jigawa: SBMCs in schools with more years of exposure to Output Steam 4 intervention are more functional

Comparison of school management by Output Stream intervention group in 2016



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 (which may have important implications for the extent to which changes in teacher competence translate into improvements in teaching practices).

4.1 Teacher competence

Box 14: Teacher competence: Key findings

- Compared to 2012, teachers in Jigawa are slightly more competent in 2016, although the difference is not statistically significant.
- In 2016, 68% of teachers are competent on a measure based on lesson observations and teachers' curriculum knowledge.
- However, only 4% of teachers meet a stricter standard of competence that includes passing primary grade-level literacy and numeracy content knowledge tests. Only 6% of teachers are able to pass both of these tests.
- There appears to be no effect of ESSPIN training on teacher competence, although it should be noted that ESSPIN-trained teachers may be sharing some of their training with their colleagues.
- Compared to 2014, teachers perform significantly worse across all domains of the English literacy and numeracy tests. The decline over time is significant for the English literacy tests.
- ESSPIN-trained teachers perform marginally better on the content knowledge tests, compared to non-ESSPIN-trained teachers.
- ESSPIN-trained teachers are more motivated than non-ESSPIN-trained teachers.

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 mathematics), two relate to teaching practices, and one 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 mathematics; see Box 15).

For CS2 and CS3, a stricter version of the competence indicator was developed. The criterion on 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 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 15: 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 mathematics. 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.

Compared to 2012, teachers in Jigawa are slightly more competent in 2016, although the difference is not statistically significant. In 2016, teachers are significantly more likely to use at least one teaching aid during their lesson, with almost all teachers doing so (99.8%). Teachers are also more likely to be using more praise than reprimands during their lessons than they were in 2012.

This is offset by fewer teachers assigning at least two individual or group tasks, compared to both 2012 and 2014. This may be an aspect of teaching that becomes increasingly difficult as class sizes increase, as has been the case in Jigawa over the past years.

Compared to 2012, fewer teachers knew the English or mathematics curriculum benchmarks, but the proportion of teachers who know the benchmarks has not changed compared to 2014. As we have noted in the CS2 report, there were some inconsistencies in how this question was administered in CS1 compared to the other rounds of the survey, which may explain the initial drop in curriculum knowledge.[7]

There are large gaps in teachers' content knowledge and this has worsened slightly since 2014. Only 6% of teachers are able to pass both a primary grade-level English and numeracy test. It is worth exploring why the SIP has not made greater progress on this front, and what could be done to address this going forward. As a result of poor performance on the content knowledge tests, only 4% of schools meet the stricter teacher competence measure. The proportion of schools meeting this measure has not changed since 2014.

[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.

Table 20: Jigawa: Teacher competence in CS1, CS2 and CS3

	2012 (CS1)	2014 (CS2)	2016 (CS3)	Change 2012–2016	Change 2014–2016
(1) Knowledge of Eng./mathematics curriculum (%)	57.1	19.0	19.0	-38.0*	0.0
(2a) Use of one or more teaching aid (%)	89.4	98.6	99.8	+10.4*	+1.2*
(2b) Use of one or more teaching aid, excl. blackboard (%)		72.7	61.3	n/a	-11.4*
(3) Praise more than reprimand (%)	63.2	83.5	91.8	+28.6*	+8.3
(4) Assigns two or more ind./group tasks (%)	47.0	46.1	27.8	-19.2*	-18.3*
(5) Passes English and mathematics test		9.7	6.1	n/a	-3.6
Teacher competence score (% of criteria fulfilled)	65.2	67.1	65.5	+0.3	-1.6
Teacher competence standard fulfilled (%)	62.2	64.5	68.6	+6.5	+4.1
Teacher competence score (% of criteria fulfilled, strict version)		55.5	50.3	n/a	-5.2*
Teacher competence standard fulfilled (strict) (%)		5.8	4.3	n/a	-1.5

Note. * indicates statistical significance ($p < .05$). The CS2 version of the competence score adds the teacher's performance in the literacy and numeracy tests to the number of other criteria met by the teacher. For example, a teacher who met all four original criteria and also scored 100% in the literacy and numeracy tests would receive a competency score of 100%.

We also examine how teachers who reported having received ESSPIN training by CS3 performed compared to those that did not report having received ESSPIN training.

Teacher competence does not differ by whether a teacher has been trained by ESSPIN or not. ESSPIN-trained teachers are somewhat more likely to assign at least two individual or group tasks during their lesson, but somewhat less likely to be using more praise than reprimands. It is worth noting 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. It is therefore possible that the lack of effect reflects the fact that ESSPIN-trained teachers have been successful at sharing some of their training with their colleagues, particularly on those aspects of teacher competence that have improved over time (using a teaching aid and using praise).

On our original measure of teacher competence, ESSPIN-trained teachers appear to be performing worse than non-ESSPIN-trained teachers, but this is influenced to some extent by the measurement of the standard. The indicator on knowledge of the English/mathematics curriculum is only factored into the measurement of the teacher competence standard for teachers who teach

either English or mathematics. ESSPIN-trained teachers are more likely to teach English or mathematics than non-ESSPIN-trained teachers. Across the board, all teachers perform poorly on knowledge of the curriculum, but this indicator is factored into the measurement of teacher competence more often for ESSPIN-trained than for non-ESSPIN-trained teachers. In light of this, we have calculated a teacher competence standard that excludes the indicator of curriculum knowledge. On this standard, ESSPIN-trained and non-ESSPIN-trained teachers perform similarly, and both are very likely to meet the standard.

On the strict measure of teacher competence, ESSPIN-trained and non-ESSPIN-trained teachers perform similarly, and very few teachers in either group are able to meet the standard. This again is driven by the poor performance on the content knowledge tests, where the proportion of teachers passing both tests did not differ by ESSPIN training. In addition, while almost all teachers are using a teaching aid, reading from, writing on or having pupils copy from the blackboard is the only teaching aid used by many teachers (39%), who fail to make use of more innovative teaching aids.

Table 21: Jigawa: Teacher competence in CS3 by intervention group

Intervention group	Non-ESSPIN-trained	ESSPIN-trained	Difference
(1) Knowledge of Eng./mathematics curriculum (%)	17.2	19.6	+2.4
(2a) Use of one or more teaching aid (%)	99.5	100.0	+0.5
(2b) Use of one or more teaching aid, excl. blackboard (%)	61.6	61.2	-0.4
(3) Praise more than reprimand (%)	95.0	90.0	-4.9
(4) Assigns two or more ind./group tasks (%)	23.5	30.1	+6.6
(5) Passes English and mathematics test	5.3	6.6	+1.2
Teacher competence score (% of criteria fulfilled)	67.4	64.5	-2.9
Teacher competence standard fulfilled (%)	78.7	63.2	-15.5
Teacher competence standard (excl. curriculum knowledge, %)	95.4	92.2	-3.2
Teacher competence score (% of criteria fulfilled, strict version)	51.9	49.4	-2.5
Teacher competence standard fulfilled (strict) (%)	3.7	4.7	+1.0
Additional indicators:			
Proportion of time spent – explaining (%)	56.3	59.6	+3.3
– instructing / presenting / dictating (%)	14.3	12.6	-1.6
– chanting (%)	8.1	6.8	-1.3
– closed question / response (%)	1.7	2.5	+0.9
– open question / response (%)	11.0	11.7	+0.7
Proportion of time spent speaking English (%)	9.4	11.0	+1.6
Teacher summarised the lesson (%)	67.5	75.9	+8.4
Teacher revisited the lesson's objectives (%)	38.0	37.6	-0.4
Teacher gave learners homework (%)	24.9	23.1	-1.8
Teacher tested learners' knowledge (%)	55.4	61.3	+5.8
Teacher marked learners' written work (%)	8.9	7.4	-1.5
Note. * indicates statistical significance (p < .05)			

4.2 Findings from the teacher content knowledge tests

The findings above suggest that teachers' content knowledge worsened slightly between 2014 and 2016. Percentage scores in the teacher content knowledge tests provide a rough indication of teachers' test performance, but analysis using IRT provides more reliable learning scales that can also be interpreted more readily in terms of learning benchmarks (see Allen, 2016a). The 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 22). For example, a teacher in Band 2 for literacy is one who shows knowledge of some basic phonics, can write a simple sentence, and perform basic comprehension of a passage, as well as satisfying 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 22: 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 a carry, 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.

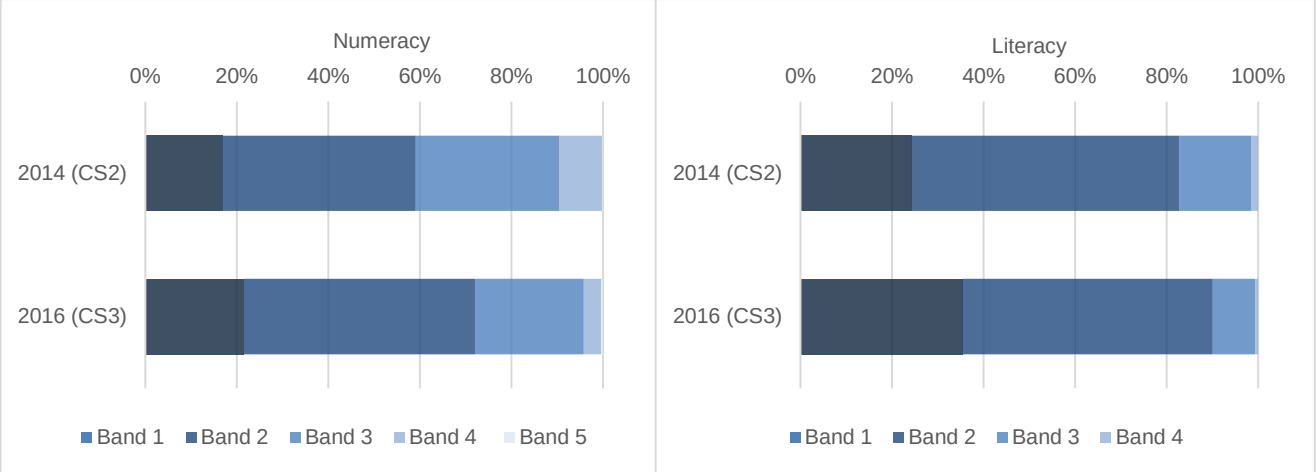
The CS3 findings point to a significant deterioration in teachers' performance since 2014 on the English literacy tests. Teachers' scale scores in English have declined by about 0.35 of a standard deviation, and this is accompanied by a larger proportion of teachers scoring in the lowest performance band (35% in 2016 compared to 24% in 2014). Performance on the numeracy tests has also declined over time, and more teachers now score in the lower performance bands, but these results are not statistically significant.

Overall, performance on the tests is poor. In 2016, over a third of teachers have only a very limited understanding of literacy concepts, and 22% have a very limited understanding of numeracy concepts. Almost no teachers are able to score in the highest performance band of a primary grade-level test.

Table 23: Jigawa: Teachers' test scores (IRT analysis) in 2014 and 2016

	2014 (CS2)	2016 (CS3)	Change on average, 2014–16
English IRT scale score (mean 500, s.d. 100)	429.8	395.3	**-34.5***
English Band 1 (%)	24.3	35.4	+11.2
English Band 2 (%)	58.5	54.6	-3.8
English Band 3 (%)	15.9	9.3	-6.5
English Band 4 (%)	1.4	0.6	-0.8
Reading (English sub-scale, mean 500, s.d. 100)	432.5	401.1	**-31.3***
Writing (English sub-scale, mean 500, s.d. 100)	443.1	411.5	**-31.5***
Grammar (maths sub-scale, mean 500, s.d. 100)	430.3	403.0	**-27.3***
Mathematics IRT scale score (mean 500, s.d. 100)	443.7	427.5	-16.2
Mathematics band 1 (%)	17.0	21.5	+4.5
Mathematics Band 2 (%)	42.1	50.6	+8.6
Mathematics Band 3 (%)	31.4	23.6	-7.8
Mathematics Band 4 (%)	9.4	3.8	-5.6
Mathematics Band 5 (%)	0.2	0.4	+0.2
Number concepts (maths sub-scale, mean 500, s.d. 100)	447.5	427.5	-20.0
Calculation (maths sub-scale, mean 500, s.d. 100)	444.2	435.8	-8.4
Note. * indicates statistical significance (p < .05)			

Figure 2: Proportion of teachers in each performance band, by year

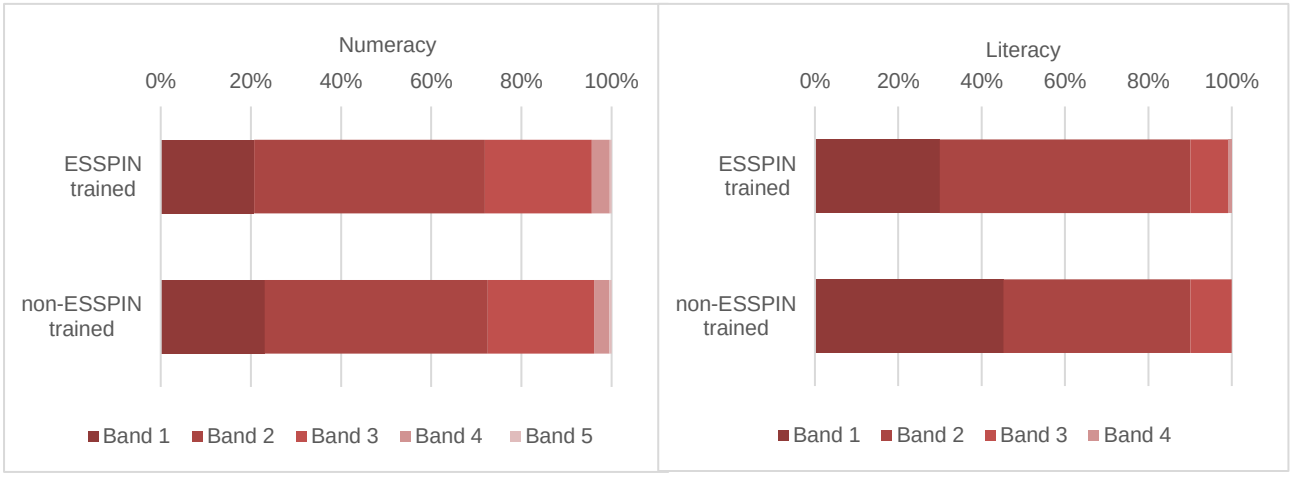


ESSPIN-trained teachers perform better on the content knowledge tests than non-ESSPIN-trained teachers, although the results do not reach the level of being statistically significant. On the English test in particular, ESSPIN-trained teachers are more likely to be scoring in the second performance band and less likely to be scoring in the lowest band compared to non-ESSPIN-trained teachers. On the numeracy tests, ESSPIN-trained teachers are more likely to be in the higher performance bands, but the differences are very small.

Table 24: Jigawa: Teachers' test scores (IRT analysis) by ESSPIN training

	Non-ESSPIN-trained	ESSPIN-trained	Difference in means
English IRT scale score (mean 500, s.d. 100)	384.2	401.5	+17.3
English Band 1 (%)	45.2	29.9	-15.3*
English Band 2 (%)	44.9	60.1	+15.2*
English Band 3 (%)	9.9	9.0	-0.8
English Band 4 (%)	0.0	0.9	+0.9
Reading (English sub-scale, mean 500, s.d. 100)	386.7	409.2	+22.5
Writing (English sub-scale, mean 500, s.d. 100)	413.3	410.5	-2.8
Grammar (maths sub-scale, mean 500, s.d. 100)	390.6	409.9	+19.3
Mathematics IRT scale score (mean 500, s.d. 100)	419.4	431.9	+12.5
Mathematics Band 1 (%)	23.1	20.7	-2.4
Mathematics Band 2 (%)	49.5	51.3	+1.8
Mathematics Band 3 (%)	23.6	23.7	+0.1
Mathematics Band 4 (%)	3.4	4.0	+0.5
Mathematics Band 5 (%)	0.5	0.4	-0.0
Number concepts (maths sub-scale, mean 500, s.d. 100)	415.6	434.0	+18.4
Calculation (maths sub-scale, mean 500, s.d. 100)	434.7	436.4	+1.6
Note. * indicates statistical significance (p < .05)			

Figure 3: Proportion of teachers in each performance band, by ESSPIN training



4.3 Teacher motivation

Teacher motivation could 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 and committed in relation to their jobs.

Box 16: 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 that had been 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 25 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 disagrees', '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 25: 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 in relation 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)		

Table 26 shows the levels of motivation among teachers, as reported during CS3, comparing those who reported having received ESSPIN training in CS2 and those that did not report having received ESSPIN training in CS2. ESSPIN-trained teachers are significantly more motivated than non-ESSPIN-trained teachers. They report feeling more engaged, more satisfied and have a higher perception of their competencies. Their perception of their colleagues' commitment and collaboration is, however, slightly less positive than for non-ESSPIN-teachers.

Table 26: Jigawa: Teacher motivation and interaction by ESSPIN training

	Non-ESSPIN-trained	ESSPIN-trained	Difference in means
Collegiality	495.3	485.1	-10.2
Satisfaction	434.8	453.5	+18.7
Skills	450.5	463.1	+12.6
Engagement	417.5	443.2	**+25.7***
Composite motivation measure	438.1	455.3	**+17.2***

Note. * indicates statistical significance ($p < .05$). All scores are normalised to have an average (mean) of 500 and a standard deviation of 100.

4.4 Summary and discussion

ESSPIN-trained teachers in Jigawa are more motivated and perform slightly better on the content knowledge tests, but are not more competent than non-ESSPIN-trained teachers. There may be several possible explanations for this. Firstly, teachers who are selected for training may already be more motivated prior to any training being received. Secondly, ESSPIN training may be having an effect on teacher motivation and engagement but this has not yet translated into changes in teacher competence, at least on those aspects we have chosen to measure. Thirdly, ESSPIN-trained teachers (and head teachers) may have been able to share some of their training with their colleagues. This could particularly be the case for those indicators of teacher competence that

have improved over time across all schools: use of a teaching aid, and use of more praise than reprimands while teaching.

The propensity of teachers to assign individual and group tasks is low, there are persisting gaps in the curriculum knowledge of many English and mathematics teachers, and while most teachers use teaching aids, far fewer teachers use them interactively. These measures of teacher competence are likely to reflect the large gaps in teachers' content knowledge. Only 6% of teachers are able to pass both a literacy and a numeracy test, and performance on these tests has worsened over time. ESSPIN-trained teachers perform only slightly better on these tests, and they are not more likely to pass both tests. If teachers have such poor content knowledge, they are likely lack the skills and understanding to effectively perform some of these tasks.

5 Trends in school quality

Box 17: School quality: Key findings

- School quality in Jigawa has improved over time. In 2016, 26% of schools meet ESSPIN's standard for a good quality school, based on having competent teachers, an effective head teacher, effective school development planning and a functional SBMC. This is a large improvement compared to 2012 when only 2% of schools were of a good quality.
- This translates into an estimated 483 more schools meeting the standard for a good quality school in Jigawa in 2016 compared to 2012, and therefore approximately 126,700 learners being educated in a better-quality school.
- However, no school in Jigawa is able to meet the strict school quality standard, which requires half the teachers in the school to have passed both content knowledge tests.
- Schools that have received more years of ESSPIN intervention score higher on measures of school quality than schools that have received fewer years of ESSPIN intervention. They are also improving faster over time than schools with fewer years of ESSPIN intervention.

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 (Box 18). 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 18. 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 introduce a second, stricter version of the standard, in which teachers must get above 50% in literacy and numeracy tests to be classed as competent (see Section 4.1 and Box 15 above).

School quality has improved in Jigawa over time. About 26% of schools meet the school quality standard in 2016, compared to only 2% of schools in 2012. While the proportion of schools meeting the standard is still low, this is nevertheless a substantial improvement. It translates into an estimated 430 more schools meeting the standard for a good quality school in Jigawa in 2016 compared to 2012. This implies that approximately 160,000 children are learning in a better-quality school[10].

No school was able to meet the stricter quality standard in 2014 or 2016. Meeting the stricter standard requires half of the teachers in each school to have passed the literacy and numeracy test; with only 6% of teachers across all schools meeting this score, most schools fail to meet the stricter quality standard purely on the basis of the teacher tests. The stricter school quality standard therefore clearly sets the bar too high to be a useful distinguisher of school quality in Jigawa's schools. However, the stricter standard does remain a valuable indicator and points to the

[10] Based on the number of schools and average school size in primary schools in Jigawa reported in the Annual School Census 2014/15 (see Table 5 and Annex A).

fact that the quality of learning is likely to be restricted when teachers lack a good grasp of primary level content.

Table 27: Jigawa: School quality in 2012–2016

	2012 (CS1)	2014 (CS2)	2016 (CS3)	Change 2012–16	Change 2014–16
Quality score (%)	36	39.2	57.2	**+21.2***	**+18.0***
School meets quality standard (%)	2.2	3.6	26.2	**+24.0***	+22.6
Quality score (strict version) (%)		36.2	53.3	n/a	**+17.2***
School meets quality standard (CS2 version) (%)		0	0	n/a	0.0
Note. * indicates statistical significance (p < .05)					

Schools with more years of ESSPIN intervention are more likely to meet the school quality standard, although the difference is not statistically significant (Table 28). Over 37% of schools in the medium intervention group meet the school quality standard, compared to only 16% of schools in the minimum intervention group.

Table 28: Jigawa: School quality by ESSPIN intervention group in 2016

	Min.	Med.	Estimated effect of one year of full intervention
Quality score (%)	52.3	62.3	+3.9
School meets quality standard (%)	15.6	37.2	+10.3
Quality score (strict version) (%)	48.7	58.2	+4.2
School meets quality standard (CS2 version) (%)	0	0	n/a
Note. * indicates statistical significance (p < .05)			

Can the differences in quality between the intervention groups be attributed to the intervention, or are they associated with differences in the schools at baseline? One way of answering this is to focus on the change over time in the different intervention groups.

Intervention schools started off at similar baseline levels of school quality in 2012. Between 2012 and 2016, schools in the medium intervention group have improved more on school quality (by 29.1 percentage points) compared to schools in the minimum intervention group (16.5 percentage points). This difference is statistically significant (Table 29). Therefore, schools with more years of ESSPIN intervention are improving faster than schools with fewer years of ESSPIN intervention – a finding which suggests that ESSPIN's intervention is contributing to improvements in school quality over time.

Table 29: Jigawa: Difference in school quality between intervention groups in change over time (2012–16)

	Intervention during 2011/12–2014/15		
	0–1 years	**2–3 years**	**Difference**
2012 (CS1)	36.9	33	-3.9
2014 (CS2)	37.9	40.6	+2.6
2016 (CS3)	53.5	62.1	+8.6
Difference (2012–2016)	16.5	29.1	**+12.5***
Note. * indicates statistical significance (p < .05)			

6 Learning outcomes

The ultimate aim of ESSPIN is to improve learning outcomes in government schools in the six states. In this chapter, we examine the trends in learning outcomes over time and differences in learning outcomes between schools that have received more or less ESSPIN intervention, and we evaluate whether effects on learning achievement can be attributed to ESSPIN.

Box 19: Learning outcomes: Key findings

- Over time, learning outcomes have improved for the Grade 2 literacy test and Grade 4 numeracy test, but have not changed for the other two tests.
- ESSPIN's intervention is associated with better performance on the Grade 4 tests but not on the Grade 2 tests.
- Even after controlling for background characteristics and pre-existing differences in test scores, learners in schools with more years of ESSPIN intervention perform significantly better on the Grade 4 tests than learners in schools with fewer years of intervention. There was no effect on Grade 2 scores.

6.1 Pupil learning achievement in English literacy and numeracy

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 (mean) of 500 and 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 30 and Table 31 list some examples of the tasks within each band.

Table 30: Examples of knowledge and skills that learners in each literacy band can demonstrate

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 the names of familiar animals 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 the names of familiar objects and animals

Table 31: Examples of knowledge and skills that learners in each numeracy band can demonstrate

Band 5: Grade 5 and 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

Learning outcomes for two of the tests have improved compared to 2014, while outcomes for the other two tests have not changed significantly. Learners' performance on the Grade 2 literacy test and the Grade 4 numeracy test has improved significantly since 2014. This is reflected in a smaller proportion of learners being found in the lowest performance band, and a higher proportion of learners being found in the higher performance bands. On the Grade 4 numeracy test, in particular, a larger proportion of learners are now found in the highest two performance bands compared to previous years. While performance on the Grade 4 numeracy test has also improved, compared to both 2012 and 2014, performance on the Grade 2 literacy test is similar to 2012. On the other two tests, learners are performing slightly worse over time, although the results are not statistically significant.

Given that learner enrolment in Jigawa has increased by 17%, even small improvements in the proportion of learners falling into higher achievement bands means that the overall number of children that are better educated has increased. Nevertheless, learners' test performance remains very poor. On the literacy tests, only between 0.9% (Grade 4) and 3% (Grade 2) of learners were able to perform at their expected grade level. In the numeracy tests, these proportion are slightly higher, with 7.2% (Grade 2) and 14.2% (Grade 4) of learners performing at the expected level or better.

Table 32: Jigawa: Learning outcomes in 2012–16

	2012 (CS1)	2014 (CS2)	2016 (CS3)	Change 2012–16	Change 2014–16
Grade 2 literacy score	433.9	402.2	424.6	-9.3	**+22.4***
– Band 0: Pre-school (%)	67.9	91.2	84.7	+16.7*	-6.5
– Band 1: Grade 1 (%)	19.2	8.4	12.3	-6.9	+3.8
– Band 2: Grade 2 (%)	12.9	0.4	3	-9.8*	+2.7*
Grade 4 literacy score	411.5	405.4	400.1	-11.4	-5.3
– Band 1: Grade 1 (%)	83	91.2	91.4	+8.4*	+0.2
– Band 2: Grade 2 (%)	5.6	4.9	5.9	+0.2	+1.0
– Band 3: Grade 3 (%)	2.7	0.4	1.8	-0.9	+1.4
– Band 4: Grade 4 (%)	8.7	3.5	0.9	-7.8*	-2.6
Grade 2 numeracy score	445.1	421	413.6	-31.5	-7.4
– Band 0: Pre-school (%)	28.1	22.3	17.6	-10.5	-4.7
– Band 1: Grade 1 (%)	52.9	71.2	75.1	+22.3*	+3.9
– Band 2: Grade 2 (%)	19	6.5	7.2	-11.8*	+0.7
Grade 4 numeracy score	405.7	385.8	429.2	+23.5	**+43.4***
– Band 1: Grade 1 (%)	42.2	62.3	46.6	+4.4	-15.8*
– Band 2: Grade 2 (%)	30.9	26.5	27.8	-3.1	+1.3
– Band 3: Grade 3 (%)	18.4	9.6	11.5	-6.9	+1.9
– Band 4: Grade 4 (%)	7.4	1.6	11.8	+4.4	+10.2*
– Band 5: Grade 5 (%)	1.1	0.1	2.4	+1.3	+2.3*

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

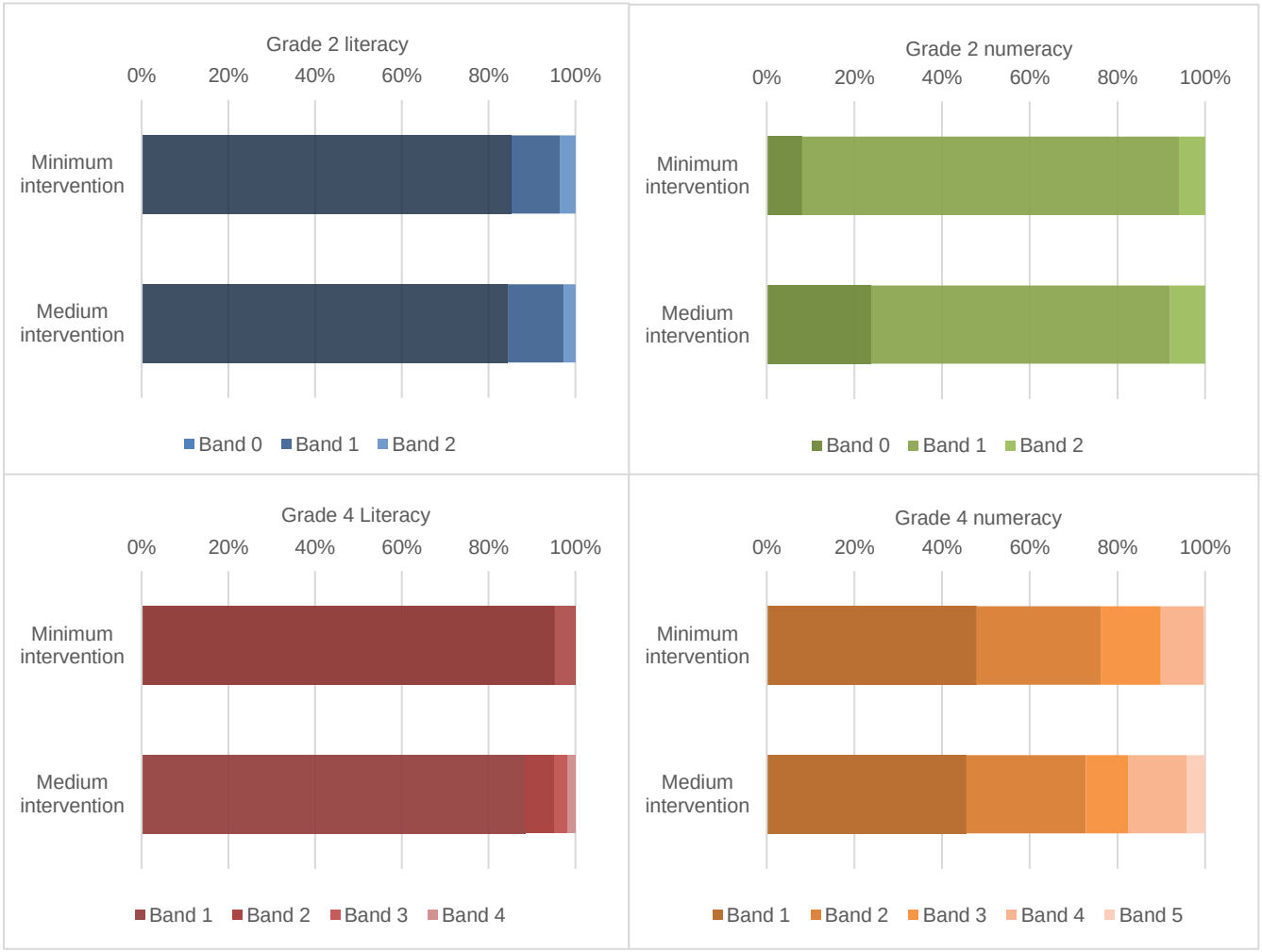
We also compare learning outcomes by amount of ESSPIN intervention received. In Grade 4, ESSPIN's intervention is associated with better performance on the literacy test, while on the numeracy test the difference in performance is positive but not statistically significant. On these tests, in schools with more years of ESSPIN intervention fewer learners are scoring in the lowest performance bands, and more are scoring in the highest performance bands. On the Grade 2 tests, learner performance does not vary much by amount of ESSPIN intervention.

Table 33: Jigawa: Learning outcomes by ESSPIN intervention group in 2016

	Min.	Med.	Estimated effect of one year of full intervention
Grade 2 literacy score	428.5	422	+2.2
Band 0: Pre-school (%)	85.1	84.4	-0.6
Band 1: Grade 1 (%)	11.3	12.9	+0.3
Band 2: Grade 2 (%)	3.6	2.7	+0.3
Grade 4 literacy score	394.7	404.5	**+17.9***
Band 1: Grade 1 (%)	95.2	88.4	-6.0*
Band 2: Grade 2 (%)	4.8	6.7	+2.4
Band 3: Grade 3 (%)	0	3.2	n/a
Band 4: Grade 4 (%)	0	1.7	n/a
Grade 2 numeracy score	434.9	399.8	-1.2
Band 0: Pre-school (%)	8	23.8	+2.7
Band 1: Grade 1 (%)	86	68.1	-5.6
Band 2: Grade 2 (%)	6	8	+2.9
Grade 4 numeracy score	418.3	438.1	+23.5
Band 1: Grade 1 (%)	47.8	45.5	-8.0
Band 2: Grade 2 (%)	28.4	27.3	+1.5
Band 3: Grade 3 (%)	13.7	9.7	-0.5
Band 4: Grade 4 (%)	9.8	13.4	+4.0
Band 5: Grade 5 (%)	0.3	4.1	+4.1
Note. * indicates statistical significance ($p < .05$)			

Figure 4 illustrates pupils' poor performance on the literacy tests, with the majority of learners falling into the poorest performance band. In Grade 4, we can see that only in the medium intervention group are a small proportion of learners beginning to score in the highest performance band.

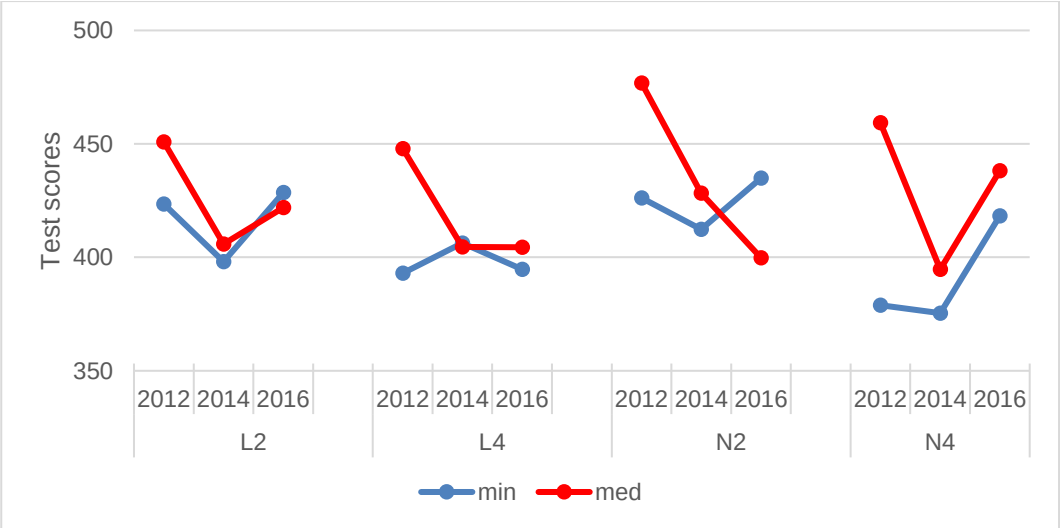
Figure 4: Jigawa: Distribution of test scores by intervention group in 2016



In Figure 5, we disaggregate the change over time according to ESSPIN intervention. Schools that received more years of ESSPIN intervention had better learning outcomes at baseline in 2012 than schools with fewer years of ESSPIN intervention. Surprisingly, the overall trend seems to suggest that schools with fewer years of ESSPIN intervention are catching up to (and in some cases overtaking) the schools with fewer years of intervention.

These patterns are explored more rigorously in Section 6.2 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: Jigawa: Learning outcomes by test, year and ESSPIN intervention



6.2 Controlling for school and pupil characteristics

6.2.1 Differences in background characteristics

As has been noted in Section 1.3 and Annex A, schools in the minimum and medium intervention groups in Jigawa differ from each other in regard several background characteristics. The schools with more years of ESSPIN intervention are more likely to be urban, have more qualified teachers and have lower PTRs. Urban schools, in particular, tend to have better learning outcomes than rural schools, and, if left uncorrected, this difference between the intervention groups could bias our estimates of ESSPIN intervention effects upwards.

There may also be differences between the intervention groups that bias our estimates of ESSPIN intervention downwards. For example, schools with more years of ESSPIN intervention have slightly worse infrastructure than schools with fewer years of intervention, and they are much larger and therefore possibly more difficult to manage.

There are a number of differences between the groups of schools that have had more years of ESSPIN intervention and those that have had fewer years of intervention, and taken together these could bias our estimates of ESSPIN's effect in either direction. We use a number of statistical methods to control for these differences in the following sections.

6.2.2 Timing of ESSPIN intervention and learning outcomes in 2016

Analysing the effects of ESSPIN's intervention is made complicated by the variations in the timing and duration of the intervention for different schools in Jigawa. The initial pilot schools in Jigawa received two years of 'full package' intervention, but then received no further year of 'full package' intervention until 2014/15. A further group of schools has consistently received intervention since 2012/2013, while the largest group of schools began receiving intervention only in 2014/15.[11]

[11] There is another group of schools which started receiving the intervention in 2013/14, and which has received two years of full intervention to date. We were not able to include this group of schools in this analysis because our sample contains only four schools that fall into this category. This number is too small to robustly detect any differences between this group and other groups.

The timing of the intervention could affect learning outcomes in at least two ways. First, it may take time for the effects of leadership training, teacher training and other interventions to filter through to measurable gains in learning outcomes. This is especially so given that children are tested in the second and fourth grades. From this perspective, for example, a Grade 4 learner whose teachers benefited from training during the current school year is unlikely to benefit as much as one whose teachers were trained four years ago, so that they have had consistent exposure to better teaching. This would suggest that training that was initiated a longer time ago would have larger effects on learning outcomes. Second, however, it is possible that the effects of training wear off over time, particularly if it is common for teachers to transfer between schools, or if trained teachers retire. More recent training would then have a larger effect on learning outcomes.

We explore this using a regression analysis that compares every combination of timing and duration of intervention (with the average across these intervention patterns serving as a baseline for comparison). In addition to a simple regression model, we report a model that also controls for differences in school characteristics (Table 34).

Schools that started to receive the intervention recently, and have therefore had a shorter duration of intervention (one to two years), have lower learning outcomes on the Grade 4 tests compared to the schools that have had three or more years of intervention. Comparing the two groups of schools that have had three years of intervention, those that received the intervention most recently had higher learning outcomes on the Grade 4 numeracy test, but not on the literacy test. When we control for background characteristics between schools, these results on the Grade 4 tests are still in the same direction, but they are no longer statistically significant. Performance on the Grade 2 tests does not vary much by timing of ESSPIN intervention. Therefore, after controlling for background characteristics, there is not enough evidence to suggest that the precise timing and duration of ESSPIN intervention has had an impact on learning outcomes.

Table 34: Jigawa: Difference in test scores in 2016 by timing of ESSPIN intervention

Years of intervention	Total years	L2		L4		N2		N4	
2014/15 and 2015/16	1	-6.4		-27.4	*	-2.0		-41.8	*
2012/13–2014/15	3	+8.2		+13.8		+4.1		+29.6	*
2009/10-10/11 and 14/15	3	-1.8		+13.6		-2.1		+12.2	
Controlling for background characteristics									
2014/15 and 2015/16	1	+8.8		-19.1		+8.3		-22.3	
2012/13–2014/15	3	+2.0		+11.7		-1.2		+19.6	
2009/10-10/11 and 14/15	3	-10.8		+7.4		-7.1		+2.6	
Note. * indicates statistical significance (p < .05)									

6.2.3 Are learning outcomes better in schools with more years of intervention in 2016?

In Section 6.1 above we found that learners from schools that have received more years of intervention had better learning outcomes in Grade 4 than those from schools that received fewer years of intervention. However, there are also some pre-existing characteristics of the schools that received more years of intervention which might have biased these results upwards (see Annex A). Therefore, in this section we add statistical controls for these 'confounding variables' – 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's intervention, conditional on school characteristics and pre-existing differences in test scores. We present four different models that control for a range of pre-existing differences between schools and learners.

We find significant positive effects of ESSPIN's intervention on Grade 4 learning outcomes. The effect appears to be large (0.3 standard deviations (SD) for literacy, 0.6 SD for numeracy), but reduces in magnitude when we control for school characteristics (to 0.2 SD for literacy and 0.3 SD for numeracy; Model 2) and in the case of literacy becomes statistically non-significant. When we focus on intervention during 2012–16, and control for learning outcomes at the beginning of this period, the estimated effects are smaller still (0.05 to 0.1 standard deviations; Models 3 and 4), though they are still statistically significant.

Across all models, learners in schools with more years of ESSPIN intervention are not performing any better on the Grade 2 tests than learners from schools with fewer years of intervention.

Table 35: Jigawa: Estimates of the effect of ESSPIN's intervention on learning outcomes in 2016

Model	L2		L4		N2		N4	
(1) Simple regression, clustered SE, no sample weights	+7.75		+38.34	*	+0.75		+60.02	*
(2) Full covariates	-16.91		+21.49		-14.77		+31.55	*
(3) Lagged school-level learning outcomes	+3.84		+6.32	*	+2.22		+10.92	*
(4) Lagged outcomes and covariates	+0.6		+5.23	*	-1.01		+6.96	*
Note. * indicates statistical significance (p < .05)								

6.3 Summary and discussion

We have applied IRT to measure learners' performance in literacy and numeracy, in Grades 2 and 4. Over time, learning outcomes improved on the Grade 2 literacy test and the Grade 4 numeracy test, while they worsened slightly, but non-significantly, on the Grade 4 literacy test and the Grade 2 numeracy test. Grade 4 learning outcomes are better in schools that have received more years of ESSPIN intervention. In Grade 2, there is no difference in learning outcomes based on ESSPIN intervention.

There are some significant pre-existing differences in the schools that have received more years of ESSPIN intervention and those that received fewer years of intervention. We used a number of statistical methods to control for these differences and to reduce bias in our estimates of the effect of ESSPIN's intervention. We also considered whether learning outcomes changed depending on the timing of the ESSPIN intervention received. Our findings on the timing of the ESSPIN intervention suggest that learning outcomes in schools with prolonged and recent intervention are better on the Grade 4 tests. However, these results are not statistically significant after controlling for school characteristics, and we therefore do not have sufficient evidence to suggest that the timing of the ESSPIN intervention is impacting learning outcomes.

We find significant positive effects of ESSPIN's intervention on Grade 4 learning outcomes, although our findings are ambiguous regarding whether the effect is small or large, depending on the model used. We cannot tell from our results exactly what differentiates Grade 2 and Grade 4 learners. In general, Grade 4 learners have been in an ESSPIN school for longer than Grade 2 learners – and this may imply that learning outcomes only change once learners have been exposed to ESSPIN-trained teachers and head teachers for a longer period of time. Some support

for this explanation comes from the CS2 report, where we found that learners in Grade 2 were not yet performing better in schools with more years of ESSPIN intervention. Although we did not re-sample the same learners, this cohort of learners is now in Grade 4 – and, as a cohort, is showing improvements in learning outcomes. However, it is also possible that learners in Grade 4 were simply taught by ESSPIN-trained teachers more often than learners in Grade 2.

7 Conclusions and implications for ESSPIN in Jigawa

The Composite Surveys' findings suggest that some progress has been made on school-level outcomes in Jigawa State. Since 2012, some indicators have improved significantly (SBMC functionality, school development planning effectiveness and overall school quality), but others have remained the same (head teacher effectiveness, school inclusiveness, inclusiveness of women and children in SBMCs). While head teachers in Jigawa are not more effective overall, they are more likely to have carried out lesson observations and professional development meetings in 2016, which was a key focus of training and was an aspect of school leadership most under their direct control.

Similarly, when one considers the level at which schools are performing as at 2016, outcomes are mixed. About 37% of schools meet the school development planning standard in 2016, 68% of SBMCs meet the functionality standard, and 26% of schools meet the standard for a good quality school. These represent very large improvements over time: they translate into approximately 252,790 more learners going to a school with a functioning SBMC in 2016, compared to 2012, while approximately 126,700 more learners attend a good quality school according to ESSPIN's standard. On the other hand, only 17% of schools meet the standard on head teacher effectiveness, 14% do so on inclusion, and no school in Jigawa meets the stricter school quality standard as a result of poor performance on the teacher tests.

The performance of schools with more years of ESSPIN intervention is mixed. Schools with more years of ESSPIN intervention are better at school development planning, have more functional SBMCs, and their SBMCs are more encouraging of children's participation. However, head teachers in schools with more years of ESSPIN intervention are no more effective than head teachers in schools with fewer years of intervention, and schools are no more inclusive. Overall, ESSPIN's intervention appears to be contributing to increases in school quality, as schools with more years of ESSPIN intervention improve faster on this indicator over time than schools with fewer years of intervention.

ESSPIN-trained teachers in Jigawa are more motivated and perform slightly better on the content knowledge tests, but are not more competent than non-ESSPIN-trained teachers. It is possible that ESSPIN-trained teachers have been able to share some of their training with their colleagues. This could particularly be the case for those indicators of teacher competence that have improved over time across all schools: use of a teaching aid, and use of more praise than reprimands while teaching. The possibility that ESSPIN training has had some effect on teachers is supported by the fact that Grade 4 learning outcomes have improved more in schools with more years of ESSPIN intervention.

However, teachers are performing very poorly on literacy and numeracy content knowledge tests, and performance has worsened compared to 2014. Most teachers are able to complete simple primary school-level tasks, such as writing a simple sentence in English or word problems for addition in mathematics, but they stumble with more advanced tasks, such as extracting basic information from a passage or using a number line to represent sums. In addition, schools continue to have serious problems with teacher attendance. Only 40% of teachers are in their classrooms on time in the mornings, and ESSPIN's intervention has not had an effect on teacher presence.

Learner performance in Jigawa remains poor, with few pupils able to perform at their expected grade level. However, performance has improved over time on the Grade 2 literacy and Grade 4 numeracy tests (while it has worsened slightly, but non-significantly, on the Grade 4 literacy and Grade 2 numeracy tests). We find significant positive effects of ESSPIN's intervention on Grade 4 learning outcomes, but no such effect on Grade 2 learning outcomes. The positive effect of

ESSPIN's intervention on Grade 4 outcomes is robust to controlling for pre-existing school characteristics and past test performance. Our results are ambiguous as to whether to classify this effect as large compared to other learning programmes (0.3 SD) or small (0.05 SD) – this depends on the model used.

These mixed findings should be interpreted in the light of contextual factors, particularly the increase in learner enrolment and PTRs that has occurred in Jigawa since 2012. This is likely to have put pressure on head masters and school systems, and to have made teaching conditions more difficult for teachers. It also means that even small improvements in the proportion of learners scoring in higher performance bands translates into greater numbers of children being better educated.

ESSPIN's intervention appears to be producing small, incremental change in schools. For example, learning outcomes improved only in the Grade 4 cohort, which has generally been exposed to ESSPIN schools for longer. These findings suggest that intensified ESSPIN intervention, such as more training for more teachers, might result in more substantial improvements in learning outcomes. On the other hand, schools face considerable constraints in terms of teachers' content knowledge and teacher attendance in the classrooms. These constraints are likely to remain key barriers to larger improvements in learning outcomes, and it is questionable whether ESSPIN's intervention on its own would be sufficient to raise learning outcomes to grade-appropriate levels within a reasonable time-frame.

Bibliography

Bennell, P., and 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 (2013b) Extension of the Education Sector Support Programme in Nigeria, August 2014 – January 2017. Business case for DFID.

Little, D. and Pinnock, H. (2014). Qualitative review of ESSPIN's support to School-Based Management Committees. Abuja, Nigeria: ESSPIN.

Megill, D. (2014a) Recommendations on Sampling Plans for the Second Composite Survey and the GEP-3 Unified Survey (GUS) in Nigeria. Unpublished note written for ESSPIN and OPM.

Megill, D. (2014b) Final sample design and weighting procedures for ESSPIN Second Composite Survey (CS2) in Nigeria. Unpublished note written for ESSPIN and OPM.

Mezger, C. (2016) Trends of inequality in access to schooling in Nigeria. Evidence from three DHS surveys (2003, 2008, 2013). Abuja: Education Data, Research and Evaluation in Nigeria (EDOREN).

Poulsen, H. (2009). School Based Management Committees (SBMC) in Policy and Practice: Research Synthesis Report (Volume 1). Abuja, Nigeria: ESSPIN.

RTI International (2014) 'Nigeria reading and access research and activity (RARA): adaptation of Education Sector Support Programme in Nigeria's teacher capacity development strategy'. Document produced for the United States Agency on International Development (USAID).

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

Annex A School characteristics

The table below sets out summary statistics for Jigawa's schools, split by categories according to the level of Output Stream 3 intervention (minimum, medium, maximum). The data come from the Annual School Censuses from 2009/10, 2013/14 and 2014/15.

Jigawa's schools by level of ESSPIN intervention	Total	Min.	Med.	
Distance from LGA headquarters	**16.3**	17.8	14.5	*
Age of the school in 2014	**33.2**	30.7	36.0	*
Urban (%)	**12.2**	7.4	17.9	*
Nomadic (%)	**6.2**	8.7	3.3	*
Islamic (%)	**4.1**	4.8	3.3	
Double shift (%)	**-**	-	-	
Had parent–teacher association in 2014/15 (%)	**95.3**	95.1	95.4	
Had SBMC in 2014/15 (%)	**96.3**	95.1	97.6	*
PTR in 2009/10	**45.5**	47.3	37.8	*
PTR in 2013/14	**59.9**	63.3	43.5	*
PTR in 2014/15	**54.8**	60.8	47.9	*
% change in PTR between 2009/10 and 2013/14	**53.8**	53.8	54.2	
% change in PTR between 2013/14 and 2014/15	**10.8**	10.1	13.9	
Number of classrooms in 2014/15	**4.8**	3.8	5.9	*
Number of teachers in 2014/15	**6.4**	4.6	8.5	*
Primary enrolment in 2009/10	**228.3**	183.0	437.9	*
Primary enrolment in 2013/14	**266.9**	218.4	506.1	*
Primary enrolment in 2014/15	**262.3**	214.8	493.1	*
% change in enrolment 2009/10–2014/15 (%)	**31.4**	31.0	33.5	
% change in enrolment 2013/14–2014/15	**4.9**	5.8	0.7	*
% of teachers with academic diploma/degree	**51.6**	46.3	57.6	*
% of teachers with PGDE, BEd or MEd	**4.1**	4.0	4.2	*
% of teachers with NCE, Grade II or equivalent	**74.9**	74.7	75.2	*
School has a power source (grid/other)	**5.0**	2.8	7.5	*
% of classrooms with enough seating	**56.1**	56.9	55.5	*
% of classrooms with a good blackboard	**72.8**	75.6	70.8	*
% of classrooms in good condition/minor repairs	**88.9**	87.8	89.8	*
School has at least one toilet (%)	**49.6**	59.6	38.2	*
Number of schools	**1,937**	1,065	872	

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 'total' column includes schools that do not have an intervention code. (3) The PTRs shown in the table are calculated as the average. PTRs for schools in the state ($\frac{\sum {}^{P_i}/_{T_i}}{N}$) and not the PTR for the state as a whole (${}^{\sum P_i}/_{\sum T_i}$).

Annex B ESSPIN Output Stream 3 interventions

The table below shows the ESSPIN Output Stream 3 interventions delivered to date in Jigawa. In order to make the variation in interventions across and within states manageable for analysis, each combination of interventions was categorised as none, minimum, medium, or maximum, according to the number of years of full intervention received before 2015/2016. Full intervention means the school received some leadership training, some teacher training, and some school visits during the year, though the amount of each may vary. The schools have been grouped as follows: minimum (zero to one years), medium (two to three years), maximum (four to five years).

	Category (years of intervention)	2009/2010			2010/2011			2011/2012			CS1	2012/2013			2013/2014			CS2	2014/2015			2015/2016			CS3
		L	T	SV	L	T	SV	L	T	SV		L	T	SV	L	T	SV		L	T	SV	L	T	SV	
Jigawa	Minimum (1)																		8	3	16	8	3	14	
	Medium (2)														6	3	9		20	6	14	3		14	
	Medium (3)											6	3	9	6	3	9		4	6	14	3		14	
	Medium (3)	5*	5*	9*	10*	5*	9*						3	9		3	9		4	6	14	3		14	

L = days of leadership training; T = days of teaching training; SV = school visits; * = pilot.

Annex C ESSPIN Output Stream 4 interventions

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

	Category	2010/2011			2011/2012			CS1	2012/2013			2013/2014			CS2	2014/2015			2015/2016			CS3
		S	P	M	S	P	M		S	P	M	S	P	M		S	P	M	S	P	M	
Jigawa	No intervention																	5				
	Post-CS1 (a)											2				2		2	2	2	1	
	Post-CS1 (b)											7		2		2	2	4	2	3	4	
	Pre-CS1	7		4	r		4		6	4*			4*	6		2	2	3	1	2	4	

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

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	1.12	2.21	0.613	408	0.001469
L2	pu_exposure	No survey weights but clustered standard errors (SEs)	11.56	1.43	0	2836	0.064225
L2	intervention_binary	Binary exposure variable	7.75	8.87	0.384	408	0.003401
L2	pu_exposure	Full covariates, survey weights	-6.44	2.54	0.013	358	0.218966
L2	pu_exposure	Full covariates, no weights	-3.91	2.68	0.148	358	0.158889
L2	intervention_binary	Full covariates	-16.91	10.33	0.105	358	0.161125
L2	pu_dexp13	Lagged school-level learning outcomes	3.84	3.22	0.235	384	0.019982
L2	pu_dexp13	Lagged outcomes and covariates	0.6	3.29	0.857	334	0.151995
L4	pu_exposure	Simple model with survey weights	6.57	1.72	0	411	0.089906
L4	pu_exposure	No survey weights but clustered SEs	6.78	0.64	0	3202	0.098513
L4	intervention_binary	Binary exposure variable	38.34	9.34	0	411	0.074138
L4	pu_exposure	Full covariates, survey weights	2.08	1.92	0.282	354	0.416264
L4	pu_exposure	Full covariates, no weights	3.28	1.6	0.043	354	0.222807
L4	intervention_binary	Full covariates	21.49	11.28	0.06	354	0.220074

L4	pu_dexp13	Lagged school-level learning outcomes	6.32	2.06	0.003	387	0.134017
L4	pu_dexp13	Lagged outcomes and covariates	5.23	2.43	0.034	338	0.213682
N2	pu_exposure	Simple model with survey weights	-0.62	3.66	0.865	406	0.000129
N2	pu_exposure	No survey weights but clustered SEs	11.62	1.47	0	2801	0.057978
N2	intervention_binary	Binary exposure variable	0.75	13.73	0.956	406	1.33E-05
N2	pu_exposure	Full covariates, survey weights	-5.79	4.27	0.178	357	0.243793
N2	pu_exposure	Full covariates, no weights	-2.57	4.48	0.567	357	0.125689
N2	intervention_binary	Full covariates	-14.77	17.54	0.402	357	0.127812
N2	pu_dexp13	Lagged school-level learning outcomes	2.22	4.09	0.589	382	0.023591
N2	pu_dexp13	Lagged outcomes and covariates	-1.01	4.32	0.816	333	0.115481
N4	pu_exposure	Simple model with survey weights	7.91	3.52	0.027	410	0.056558
N4	pu_exposure	No survey weights but clustered SEs	6.9	0.7	0	3177	0.085963
N4	intervention_binary	Binary exposure variable	60.02	13.96	0	410	0.081672
N4	pu_exposure	Full covariates, survey weights	3.57	2.53	0.161	358	0.3735

N4	pu_exposure	Full covariates, no weights	3.05	2.12	0.152	358	0.270215
N4	intervention_binary	Full covariates	31.55	15.15	0.04	358	0.278373
N4	pu_dexp13	Lagged school-level learning outcomes	10.92	3.02	0	386	0.103667
N4	pu_dexp13	Lagged outcomes and covariates	6.96	3.46	0.047	342	0.276806