



Oxford Policy Management

ESSPIN Composite Survey 2

Kwara State report

Stuart Cameron and Nicola Ruddle

June 2015

Acknowledgements

The authors are very grateful to a number of people who helped in producing this note. Ravi Somani, Sourovi De and Peter-Sam Hill worked on parts of the data analysis. David Megill reviewed weights from the first Composite Survey and produced the sampling weights application for the second Composite Survey. Allan Findlay provided support with school identification and the annual school census. From ESSPIN, Jake Ross, Kayode Sanni, Fatima Aboki, Sandra Graham, Pius Elumeze, Lilian Breakell, John Kay and Simon Thomson reviewed preliminary notes and drafts and provided valuable comments. Gratitude is also due to everyone who conducted the second Composite Survey, including the staff of Oxford Policy Management and ESSPIN, state coordinators, and the staff of the State Universal Basic Education Boards, who did the data collection; and not least to the large number of headteachers, teachers and students who took the time to participate in the study.

This survey exercise 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 project manager is Stuart Cameron. For further information contact stuart.cameron@opml.co.uk.

Oxford Policy Management Limited

6 St Aldates Courtyard
38 St Aldates
Oxford OX1 1BN
United Kingdom

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

Registered in England: 3122495

Executive summary

This report presents the findings for Kwara State from the first and second rounds of the Education Sector Support Programme in Nigeria (ESSPIN) Composite Survey (CS1 and CS2), conducted in 2012 and 2014, respectively. The survey covered a wide range of indicators at the teacher, headteacher, school-based management committee (SBMC), and pupil levels, in an attempt to understand how schools in Kwara are changing over time and whether schools which receive ESSPIN interventions are working better than those which do not.

In Kwara the ESSPIN programme began in 2009/10 and covered all government primary schools in the state. All schools received teacher and headteacher training from the beginning (Output stream 3) and the SBMC and participation training (Output stream 4) was rolled out more selectively. This means that results can be analysed according to the level of Output stream 4 intervention that a school received.

The main findings are as follows:

Teacher competence: ESSPIN provided direct teacher training during 2009/10 and 2010/11, and again in 2013/14. Across the state as a whole, around 74% of teachers met standards for teacher competence, based on their knowledge of the English and mathematics curriculum and on observations of their lessons. There was no significant change in teacher competence between 2012 and 2014. ESSPIN training was delivered to some of the teachers in every school in Kwara. We do not find statistically significant differences in competence between teachers who individually received training and those who did not. In tests, teachers on average were able to answer correctly half of grade 5 level mathematics questions, and a quarter of grade 5 level English questions.

Headteacher effectiveness: In Kwara 22% of headteachers met the overall standard for effectiveness in CS2, not significantly different from the CS1 figure (19%). Most aspects of how headteachers worked had not changed between CS1 and CS2.

School development planning: There was a significant improvement in school development planning in Kwara between CS1 and CS2. A large majority of schools (80%) had school development plans (SDPs) in place in 2014, although it remained the case that relatively few could present evidence of having acted on them. The number of schools meeting the overall standard for effective school development planning rose from 6% to 25%.

School inclusiveness: We measure inclusiveness in terms of activities to improve children's access and attendance, and the extent to which teachers encourage different groups to participate during lessons. In Kwara, schools became overall less inclusive between 2012 and 2014: the proportion of schools meeting the standard fell from 45% to 20%. Schools which had received more Output stream 4 intervention were significantly more likely to meet the inclusiveness standard (46%) than schools with little or no Output 4 intervention (16%). The evidence suggests that both these groups of schools worsened at the same rate between 2012 and 2014.

SBMCs: Overall the number of schools meeting the SBMC functionality standard more than doubled from 30% to 75% between CS1 and CS2. Schools which received Output stream 4 support had significantly more functional SBMCs than those without support, and both groups improved rapidly between 2012 and 2014. SBMCs were more likely in 2014 than in 2012 to have raised issues of children's exclusion with the school, but less likely to have taken action on commonly excluded groups. Schools with Output 4 ESSPIN support were much more likely to have met either of these criteria than schools without support. Women's and children's participation in

SBMCs also rose rapidly between 2012 and 2014; again, schools with Output 4 support were doing significantly better than those without.

School quality: Comparison of school quality between CS1 and CS2 suggests that there was a significant improvement in Kwara, with the proportion of schools meeting the standard increasing from 11% to 29%. Within CS2, schools which had received more support under ESSPIN Output stream 4 tended to have a higher quality score than other schools. Both types of school improved by approximately the same rate between 2012 and 2014 regardless of the level of intervention they received.

Pupil learning: From CS1 to CS2 the average pupil test scores in Kwara improved significantly in literacy, by over six percentage points in both grades 2 and 4. However, the scores for numeracy in grade 2 worsened, from 61% down to 52%. Average test scores were roughly the same in schools which had received more Output stream 4 intervention as in schools with no Output stream 4 support. The rate of change between 2012 and 2014 did not vary with the amount of Output stream 4 intervention the school had benefited from in most subjects, except for grade 2 numeracy, in which schools which had received more support actually improved at a slower rate than schools with no support.

Overall there are quite mixed results for Kwara when comparing the indicators in 2014 with those in 2012. It seems the outputs targeted by ESSPIN Output stream 3 – teachers and headteachers specifically – saw little change between 2012 and 2014, and perhaps as a knock-on effect there was little change in pupil learning. This may be unsurprising given that the teacher and headteacher training activities occurred prior to 2012. Schools which have been receiving Output 4 intervention since 2012 have more functional and inclusive SBMCs than other schools in the state. Between 2012 and 2014 there have been large improvements in how well SBMCs function and how inclusive they are. These improvements have been state-wide and not limited to the schools receiving the most school-level intervention from ESSPIN.

Box 1. The good and bad news from the Composite Surveys in Kwara

Positive results in this report include:

- School quality increased between 2012 and 2014, with 11% of schools meeting the quality standard in 2012 and 29% in 2014.
- From CS1 to CS2 the average pupil test scores in Kwara improved significantly in literacy, by over six percentage points in both grades 2 and 4.
- The number of schools meeting the school development planning increased from 6% in 2012 to 25% in 2014.
- The more ESSPIN support that a school received for community participation, the more functioning and inclusive its SBMC tended to be in 2014.
- Across the state, schools were much more likely to have functioning SBMCs in 2014 than in 2012, with the proportion meeting the standard increasing from 30% to 70% in two years. SBMCs also became more inclusive of women and children in this time.

Some challenges identified in the report include:

- There was no significant change in the proportion of teachers meeting the competence standard between 2012 and 2014. Teachers who had received direct training from ESSPIN were no more competent than those who had not.
- Headteacher effectiveness and leadership did not improve between 2012 and 2014.
- Schools became less inclusive over the two-year period, with the proportion meeting the standard falling from 45% to 20%.
- Average test scores for pupils fell in grade 2 numeracy between 2012 and 2014.

Table of contents

| | |
|-------------------------------------------------------------------------------|------|
| Acknowledgements | i |
| Executive summary | ii |
| List of tables, figures and boxes | vi |
| List of abbreviations | viii |
| 1 Introduction | 1 |
| 1.1 ESSPIN's SIP | 1 |
| 1.2 ESSPIN in Kwara State | 2 |
| 1.3 Conflict in Kwara | 3 |
| 2 Methodology and analysis | 5 |
| 2.2 Sampling, coverage and weights | 6 |
| 2.3 Fieldwork and instruments | 7 |
| 3 Findings | 9 |
| 3.1 Teacher competence | 9 |
| 3.1.1 Main analysis | 9 |
| 3.1.2 Findings from the teacher content knowledge tests | 12 |
| 3.2 Headteacher effectiveness | 12 |
| 3.3 School development planning | 13 |
| 3.4 School inclusiveness: meeting the needs of all pupils | 14 |
| 3.5 SBMC functionality and inclusiveness | 16 |
| 3.5.1 SBMC functionality | 17 |
| 3.5.2 Women's inclusiveness | 19 |
| 3.5.3 Children's inclusiveness | 21 |
| 3.6 School quality | 23 |
| 3.7 Pupil learning achievement in English literacy and numeracy | 25 |
| 3.7.1 Main analysis | 25 |
| 3.7.2 Distribution of test scores and sub-scale scores | 28 |
| 4 Conclusion and implications of survey findings for Kwara's ESSPIN programme | 30 |
| References | 31 |
| Annex A ESSPIN Output stream 3 interventions | 32 |
| Annex B ESSPIN Output stream 4 interventions | 33 |
| Annex C Difference in differences analysis using regression | 34 |
| School inclusiveness | 34 |
| SBMC functionality and inclusiveness | 35 |
| School quality | 36 |
| Pupil learning | 36 |

List of tables, figures and boxes

| | |
|----------------------------------------------------------------------------------------------------------------------------------------|----|
| Table 1. Proportion of schools receiving the full package of ESSPIN Output stream 3 interventions | 3 |
| Table 2. Number of schools and enrolment in the 2009 and 2013 school censuses..... | 3 |
| Table 3. Kwara: Political violence: Incidents and fatalities, 2010–2014 | 4 |
| Table 4. Sample in CS1 and CS2 and population of schools, by intervention group..... | 7 |
| Table 5. Kwara: Sample coverage in CS2..... | 7 |
| Table 6. Kwara: Teacher competence in CS1 and CS2 | 10 |
| Table 7. Kwara: Teacher competence in CS2, by intervention group | 11 |
| Table 8. Kwara: Teacher competence difference in differences (comparison of means) | 11 |
| Table 9. Kwara: Headteacher effectiveness in CS1 and CS2..... | 13 |
| Table 10. Kwara: SDP effectiveness in CS1 and CS2..... | 14 |
| Table 11. Kwara: School inclusiveness in CS1 and CS2..... | 15 |
| Table 12. Kwara: School inclusiveness in CS2, by Output stream 4 intervention gro | 16 |
| Table 14. Kwara: School inclusiveness difference in differences (comparison of means)..... | 16 |
| Table 15. Kwara: Sample size of schools with SBMCs..... | 17 |
| Table 16. Kwara: SBMC functionality in CS1 and CS2..... | 18 |
| Table 17. Kwara: SBMC functionality in CS2, by Output stream 4 intervention group | 19 |
| Table 18. Kwara: SBMC functionality difference in differences (comparison of means) | 19 |
| Table 19. Kwara: SBMC's women's inclusiveness in CS1 and CS2 | 20 |
| Table 20. Kwara: SBMC women's inclusiveness in CS2, by Output stream 4 intervention group .. | 20 |
| Table 21. Kwara: SBMC women's inclusiveness difference in differences (comparison of means) | 21 |
| Table 22. Kwara: SBMC children's inclusiveness in CS1 and CS2..... | 22 |
| Table 23. Kwara: SBMC children's inclusiveness in CS2, by Output stream 4 intervention group . | 22 |
| Table 24. Kwara: SBMC children's inclusiveness difference in differences (comparison of means) | 23 |
| | 23 |
| Table 25. Kwara: School quality in CS1 and CS2..... | 24 |
| Table 26. Kwara: School quality in CS2, by Output stream 4 intervention group | 24 |
| Table 27. Kwara: School quality difference in differences (comparison of means) | 24 |
| Table 28. Kwara: Test scores and proportion of children reaching logframe indicator in CS1 and CS2..... | 26 |
| Table 29. Kwara: Test scores and proportion of children reaching logframe indicator in CS2, by Output stream 4 intervention group..... | 26 |
| Table 30. Kwara: Pupil test score difference in differences (comparison of means) | 27 |
| Table 31. Kwara: School inclusiveness difference in differences (regression) | 34 |
| Table 32. Kwara: SBMC functionality difference in differences (regression)..... | 35 |
| Table 33. Kwara: SBMC women's inclusiveness difference in differences (regression)..... | 35 |
| Table 34. Kwara: SBMC children's inclusiveness difference in differences (regression)..... | 36 |
| Table 35. Kwara: School quality difference in differences (regression with continuous intervention variable)..... | 36 |
| Table 36. Kwara: Pupil test score difference in differences (regression) | 37 |
| | |
| Figure 1. Incidents of political violence in Nigeria and Kwara | 4 |
| Figure 2. Kwara: Teacher test scores across domains of learning | 12 |
| Figure 3. Kwara: Teacher test scores by grade..... | 12 |
| Figure 4. Kwara: School quality in 2012 and 2014, by level of Output stream 4 intervention | 25 |
| Figure 5. Kwara: Pupil test scores in schools by level of Output stream 4 intervention, 2012 and 2014..... | 27 |
| Figure 6. Kwara: Distribution of pupil test scores | 28 |
| Figure 7. Kwara: Average scores in test sub-scales, CS1 and CS2 | 29 |
| | |
| Box 1. The good and bad news from the Composite Surveys in Kwara | iv |
| Box 2. Difference in differences | 6 |
| Box 3. How to interpret the analysis, and expected results | 9 |
| Box 4. Logframe standard for teacher competence..... | 10 |

| | |
|----------------------------------------------------------------------------------|----|
| Box 5. Logframe standard for headteacher effectiveness..... | 13 |
| Box 6. Logframe standard for effective school development planning | 14 |
| Box 7. Standard for school inclusiveness (meeting the needs of all pupils) | 14 |
| Box 8. Logframe standard for SBMC functionality | 18 |
| Box 9. Logframe standard for SBMC women's inclusiveness | 20 |
| Box 10. Logframe standard for SBMC children's inclusiveness | 22 |
| Box 11. Logframe standard for school quality | 23 |

List of abbreviations

| | |
|--------|-----------------------------------------------|
| ACLED | Armed Conflict Location & Event Data Project |
| CBO | Community-Based Organisation |
| CS1 | Composite Survey 1 |
| CS2 | Composite Survey 2 |
| ESSPIN | Education Sector Support Programme in Nigeria |
| LGEA | Local Government Education Authority |
| L2 | grade 2 literacy test |
| L4 | grade 4 literacy test |
| N2 | grade 2 numeracy test |
| N4 | grade 4 numeracy test |
| SBMC | School-based management committee |
| SDP | School development plan |
| SIP | School Improvement Programme |

1 Introduction

The aims of the ESSPIN Composite Surveys are to assess the effects of ESSPIN's integrated School Improvement Programme (SIP) and report on quality of education in the six ESSPIN-supported states. This report focuses on the key findings for Kwara State. The surveys address five output indicators: teacher competence, headteacher effectiveness, school development planning, SBMC functionality, and inclusive practices in schools. It also addresses one outcome indicator – school quality – and one impact indicator – pupil learning achievement.

The second round of the Composite Survey (CS2), conducted in 2014, aimed to provide post-intervention data which could be compared to data from the first round of the survey (CS1) collected in 2012, in order to evaluate the extent of improvements in key indicators and gauge programme success. A further survey will be conducted in 2016 to again assess the impact of the interventions.

This report, focusing on Kwara State, presents findings from CS2 and comparisons between CS1 and CS2, covering all of ESSPIN's output, outcome and impact indicators.

1.1 ESSPIN's SIP

The ESSPIN programme aims to bring about better learning outcomes for children of basic education school age in six states, with a range of activities at the national, state, local and school levels. It has four output streams, focusing on: (i) strengthening federal government systems; (ii) increasing the capability of state and local governments for the governance and management of schools; (iii) strengthening the capability of primary schools to provide improved learning outcomes; and (iv) improving inclusion policies and practices in basic education (ESSPIN, 2013c).

Under the third of these output streams, ESSPIN's SIP aims to provide and support the use of structured materials that ensure teachers can deliver high-quality instruction, to strengthen teachers' own understanding of literacy and numeracy concepts, and to improve academic leadership and school improvement planning by headteachers (USAID, 2014). It 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 maintain and continue improving quality gains.

Under the fourth output stream, ESSPIN aims to improve inclusion practices and to strengthen community engagement in school improvement and wider access. In particular, Output stream 4 seeks to ensure that community members, including women and girls, influence the way schools are run; that community and government organisations are better able to press for school improvement; and that schools and communities ensure that the needs of all children are met. These interventions to improve community participation through functioning SBMCs come within a challenging sociocultural context. Qualitative research by ESSPIN (2009) found that community members were often not aware of SBMCs. SBMCs also lacked clarity on their roles and responsibilities, and lacked the resources to contribute effectively to school management. It was particularly difficult for women and students to participate, as this was a cultural taboo in many areas. The programme's theory of change assumes that the interventions will improve five pillars (or outputs) of school quality: headteacher effectiveness, teacher competence, adoption of inclusive practices to meet the needs of pupils, introduction of SDPs, and establishment of functional school-based management. These pillars collectively contribute to an improvement in overall school quality (outcome), and this in turn increases pupil learning outcomes (impact).

1.2 ESSPIN in Kwara State

In Kwara the ESSPIN intervention began in 2009/10 and covered all 1,485 government primary schools in the state. The decision was made to roll out ESSPIN to all government primary schools from the beginning, and therefore there is no control group (ESSPIN, 2013a). Initially all schools received headteacher and teacher development interventions under ESSPIN Output stream 3 (see Annex A and Table 1). As part of ESSPIN Output stream 4, the level of SBMC training, women and children participation training, and mentoring visits varied across schools in Kwara; this is shown in Annex B. Key aspects of the ESSPIN programme are as follows (ESSPIN, 2013b):

- Primary school headteachers received training on:
 - academic leadership;
 - school planning;
 - management of teachers; and
 - working with the community.
- Teachers (an average of four teachers in every primary school worked with) received training on:
 - basic literacy teaching (initial reading skills);
 - basic numeracy teaching (number concepts, addition and subtraction); and
 - use of teaching aids, classroom organisation and praise.
- Some schools received two school grants in consecutive years at an average of N150,000 per year (the exact sum depending on school size) to be spent on activities agreed by the headteacher and SBMC as priorities for school improvement and included in the SDP based on a school self-evaluation.
- State School Improvement Team members (government officers, including lecturers and teachers) received ongoing training and support over a three-year period to develop structured lesson plans, and to develop the capacity of the Advisory Service Unit and School Support Officers to enable them to lead the school improvement process at school level.
- Civil society members and government officers from the department of Social Mobilisation attended training of trainers workshops to enable them to work in partnership to activate, train and mentor SBMCs across a period of approximately 24 months. A wide range of other education stakeholders, including SBMC State Task Teams, participated in these workshops.
- Community members received 16 days of training by Civil Society and Government Partnership on establishing an SBMC, which covered:
 - roles and responsibilities of an SBMC;
 - school planning and management;
 - communication and leadership;
 - change and relationships management;
 - the participation of women and children in school improvement and education decision-making;
 - resource mobilisation and financial processes;
 - child protection and participation; and
 - inclusive education and gender.

Table 1. Proportion of schools receiving the full package of ESSPIN Output stream 3 interventions

| % | 2009/10 | 2010/11 | 2011/12 | 2012/13 | 2013/14 | Any year |
|-------|---------|---------|---------|---------|---------|----------|
| Kwara | 100 | 100 | 0 | 0 | 100 | 100 |

The total number of government primary schools in Kwara fell by just over 2% between 2009 and 2013 (Table 2). At the same time, enrolment also fell marginally, and is just under 200,000. Pupil–teacher ratios are low in Kwara, at around 17 pupils per teacher overall. Between 2009 and 2013, PTRs increased throughout the state, but most rapidly in schools which received the most ESSPIN intervention under the Output 4 stream. This increase in PTR may have affected the ability to maintain quality among this group of schools, although the ratios remain low on average. By contrast, pupil–classroom ratios have dropped from around 34:1 in 2009 to 24:1 in 2013, with little difference between schools receiving more ESSPIN Output 4 intervention and those receiving less.

Table 2. Number of schools and enrolment in the 2009 and 2013 school censuses

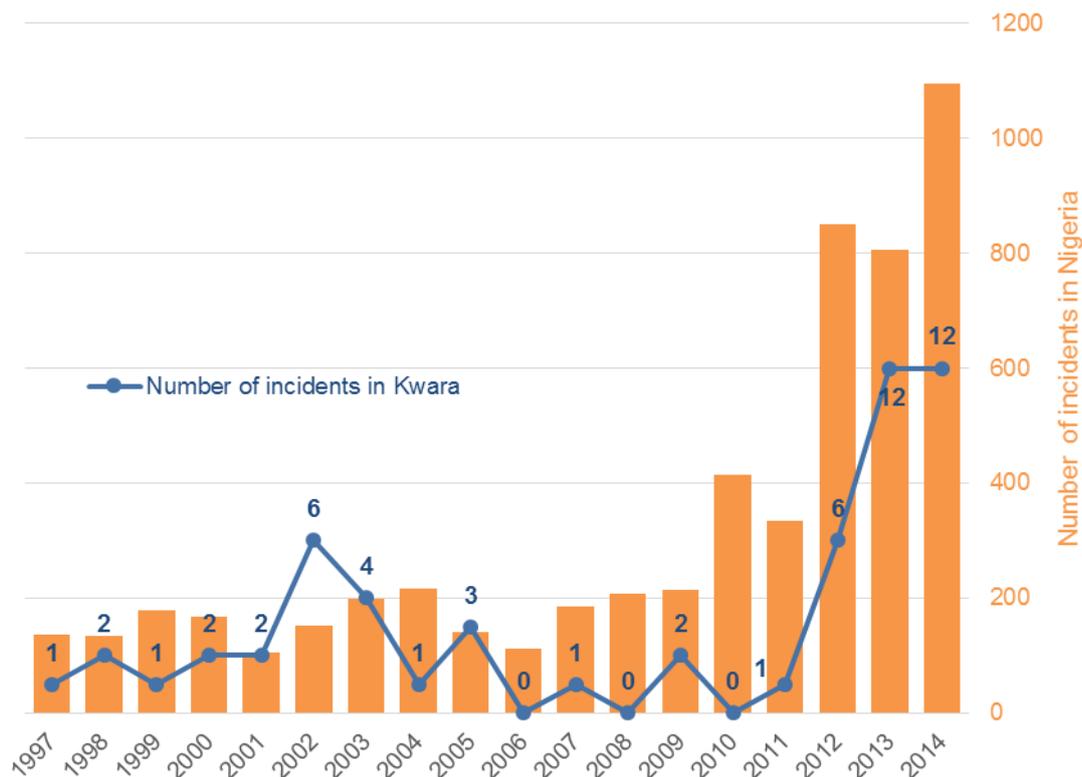
| State | 2009 | | 2013 | | Enrolment change (%) | Enrolment change (schools found in both censuses only, %) |
|-------|---------|-----------|---------|-----------|----------------------|-----------------------------------------------------------|
| | Schools | Enrolment | Schools | Enrolment | | |
| Kwara | 1,448 | 199,604 | 1,497 | 198,248 | -0.7 | -2.2 |

Note: Enrolment is for primary grades 1 to 6.

1.3 Conflict in Kwara

This report is written in the context of growing insecurity in Nigeria, particularly in but not limited to three states of the north-east in which a state of emergency has been declared (Borno, Yobe and Adamawa). Across Nigeria the number of recorded incidents of political violence and conflict has increased to eight times its 1997 level, and in Kwara it has increased from one recorded incident in 1997 to 12 in 2014 (Figure 1). There were 12 recorded violent events in both 2013 and 2014, causing 26 fatalities in 2013 and 11 in 2014 (Table 3).

Figure 1. Incidents of political violence in Nigeria and Kwara



Source: Armed Conflict Location & Event Data Project (ACLED), Version 5 (1997–2014). Note all events from ACLED are included except for those categorised as protests which did not involve a fatality.

Table 3. Kwara: Political violence: Incidents and fatalities, 2010–2014

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

Source: ACLED, Version 5 (1997–2014). Note all events from ACLED are included except for those categorised as protests which did not involve a fatality.

While in comparison with the conflict in the northern states, Kwara is not experiencing a severe situation, the level of violence is still worth bearing in mind in the reading of this report: forms of violence are likely to have some impact on education, through pupil and teacher attendance and concentration and the ability of SBMCs to meet and support school management.

2 Methodology and analysis

2.1 Evaluation strategy

2.1.1 Classifying the amount of ESSPIN intervention

The original evaluation design for ESSPIN relied on maintaining a control group of schools 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 those where intervention started more recently (Phase 2: roll-out in 2012/13 or 2013/14). In practice, the roll-out has been more complicated. In Kwara all schools joined the pilot phase in 2009/10, and have received the same level of support under Output stream 3 in each year (see Annex A for details). So this report is not able to comment on the effect of Output 3 interventions in Kwara; there is no group which received no Output 3 interventions, so there is no comparison we can use to understand what would have happened without the intervention. An exception is for teachers: among teachers in each school, only some have received direct ESSPIN training, while others have received training from their headteachers. We compare these two groups of teachers to examine the effect of personally receiving ESSPIN training versus simply being in a school where ESSPIN is in operation.

All schools received Output 3 interventions in Kwara, but not all schools received Output 4 interventions. So we can compare schools receiving little or no Output 4 intervention to those which have received full Output 4 intervention since 2011/12 (see Annex B for full details).¹ Schools that received support under Output 4 only in 2013/14 are counted with the control group, as the intervention is considered too recent to have had an effect by the time of our survey (towards the end of the same school year).

2.1.2 Modes of analysis

The purpose of CS2 is both to provide insight into the changes over time in the six states where ESSPIN works, and to evaluate whether the ESSPIN model is having an effect in the specific schools where its school improvement and community inclusion interventions have operated. We are interested in a wide range of output indicators: teacher competence, headteacher 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 in grades 2 and 4. For each of these indicators, we present in the following chapter three main types of analysis:

1. Change over time between CS1 and CS2, for Kwara as a whole. These changes likely reflect changes that are beyond the control of ESSPIN.

2. Differences between ESSPIN and non-ESSPIN schools within the CS2 results. In the group of schools that are 'expected to be better' we hypothesise that our SBMC and inclusiveness outputs, outcome and impact measures will all be higher than in the control group. If this is the case, it provides good initial evidence that ESSPIN is effective, although it does not rule out the

¹ A companion report, 'Composite Survey 2: Gender and Inclusion Report' (De and Cameron, 2015), focuses on ESSPIN's Output 4 interventions, which run in parallel with Output 3 and aim to improve inclusion and community participation in schools.

possibility that ESSPIN schools' better results could come from differences in school background characteristics pre-dating the ESSPIN intervention.

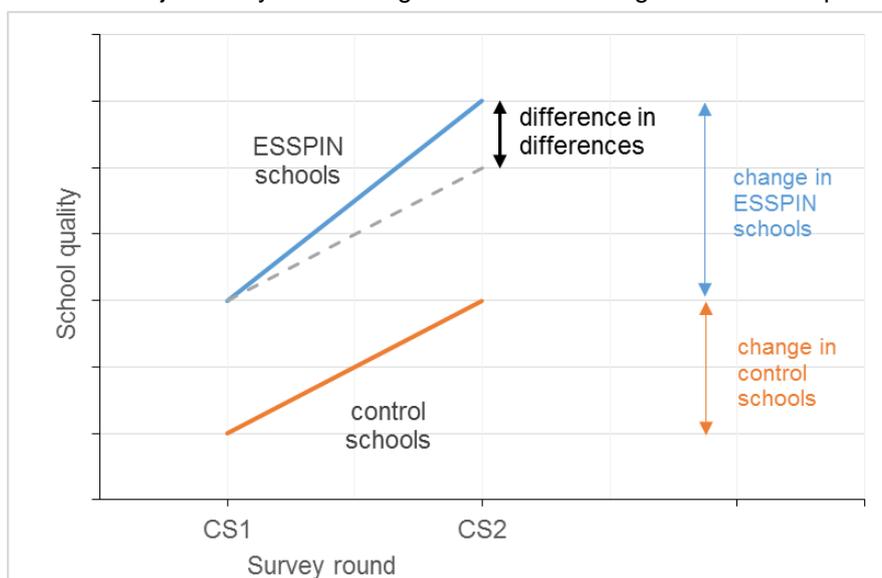
3. Difference in differences between ESSPIN and non-ESSPIN schools and over CS1 and CS2. See Box 2 below.

In each case we use statistical significance tests (t-tests or z-tests) to give an indication of whether a difference in results (over time or between intervention groups) is significant. This should not be taken as rigorous hypothesis testing (given the very large number of indicators tested), but it provides a guide to whether a difference between the weighted average results in two groups is large enough relative to the variance of the results to be able to provide us with a useful indication of likely differences in the population of schools in the six states. For analysing difference in differences we also use regression analysis; these are reported in Annex C.

Box 2. Difference in differences

The Composite Survey may reveal that ESSPIN schools are of higher quality, or have better learning outcomes, than other schools. But how do we know whether this can be attributed to ESSPIN and is not just because the ESSPIN schools were better in the first place? One way is to focus on change over time using 'difference in differences' methods. The underlying idea is that schools which have had ESSPIN interventions between CS1 and CS2 – that is, between 2012 and 2014 – ought to have improved faster during that period than schools which did not have ESSPIN interventions.

We can measure this degree of improvement by comparing averages of the indicator of interest – school quality, say – during CS1 and CS2, in the schools that received continuous Output 4 intervention to the schools that received no or minimal Output 4 intervention. Is the change over time greater in the Output 4 schools than in other schools? If so – and if statistical tests confirm that this result is unlikely to have occurred by chance – then this is considered good evidence that the Output 4 intervention itself had an effect, and that the programme was not just lucky in choosing schools that were good in the first place (selection bias).



Does a significant difference in differences (or treatment effect) prove that the faster improvement in some schools can be attributed to ESSPIN Output 4? Not absolutely. It is still possible that there are other factors at play causing a faster improvement in some schools than others. For this reason, in the overall CS2 report we use other statistical techniques to examine whether schools receiving more ESSPIN intervention had different characteristics to start with, and to control for any such differences.

2.2 Sampling, coverage and weights

In Kwara the sample allocation for CS1 was 105 schools, all of which had received some intervention. In total, 102 were able to be included in the survey, and these same schools were

visited for CS2, along with three additional schools to bring the sample size back up to 105 (Megill, 2014).

The number of schools sampled is shown in Table 4.

Table 4. Sample in CS1 and CS2 and population of schools, by intervention group

| | Category for sampling purposes | CS1 sample | CS2 sample | Population | Categories for analysis | |
|-------|--------------------------------|------------|------------|------------|------------------------------|------------------------------------------|
| | | | | | Expected to be better at CS2 | Expected to have improved during CS1–CS2 |
| Kwara | medium | 102 | 105 | 1485 | yes | no |

Note: The sample size shown is the actual sample for which data was collected.

In each school the headteacher was interviewed, as was the SBMC chairperson.

Teachers within each school were sampled from the population present in the school on the day of the survey visit and who taught grades 1–6 in the present term, using the school's teacher attendance register. The sample was reduced from 10 teachers in CS1 to six teachers in CS2 in order to improve accuracy of the indicators.

Pupils were sampled from the pupil registers for grade 2 and 4 classes – four each for numeracy and literacy by grade.

Within the schools, it was not always possible to administer all of the intended instruments. This could happen because the school was very small, and lacked a sufficient number of pupils and eligible teachers. It also sometimes happened that teachers and pupils were not present at 8am, when sampling was conducted; and occasionally pupils and teachers would leave the school after being sampled (for example, due to illness). In total 93% of the intended sample of pupils was included, and 77% of teachers. The actual numbers of schools, teachers and students sampled is given in Table 5.

Table 5. Kwara: Sample coverage in CS2

| | Schools | | Teachers | | | Pupil tests | | | |
|-------|-----------------|--------|-----------|------------|-------|-------------|-----|-----|-----|
| | Intended sample | Actual | Interview | Less. Obs. | Tests | L2 | L4 | N2 | N4 |
| Kwara | 105 | 105 | 545 | 538 | 483 | 394 | 393 | 397 | 391 |

Note: 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.

Where the schools are categorised according to the level of Output stream 4 intervention, simple averages of the results from the Composite Survey data would not be representative of what is happening across the state. This is because the profile of schools in the survey is not identical to the profile of schools in the state as a whole, in terms of the proportion of schools in each of the roll-out phases. We overcome this by applying sample weights which give greater weight to the results in schools that are relatively under-represented in the survey. Sample weights were calculated for the CS1 and CS2 schools, teachers and pupils.

2.3 Fieldwork and instruments

Fieldwork for CS2, including the pupil tests, was conducted during May and July 2014. The following data collection was carried out:

- (i) Structured interviews were conducted with teachers, headteachers, and SBMC chairpersons;
- (ii) A lesson observation was conducted for each teacher sampled;
- (iii) Teacher tests were conducted at the end of the survey, in a number of testing centres in each state; and
- (iv) Pupils in primary grade 2 and grade 4 were given tests in either literacy or numeracy.

Two indicators of aggregate learning outcomes are used in this note. The first is the total mark achieved by the pupil in each test paper, expressed as a percentage score. The second is the proportion of tested pupils who successfully answer a subset of questions which aim to measure a specific field of learning, as described in ESSPIN's logframe. Although the latter may be important for assessing ESSPIN's success in improving specific types of learning (e.g. the ability to read with comprehension), their reliance on data from a small number of questions (2–3) is statistically problematic. They are less reliable and less sensitive indicators than the total mark, which uses all the data available. For completeness, both types of indicator are used in this report.

3 Findings

Box 3. How to interpret the analysis, and expected results

For each indicator, where possible, three types of analysis are presented:

- Comparison of averages between CS1 and CS2. Here the results are representative of all schools (or teachers, or pupils) in the state, as found in CS1 and then in CS2. This depends on both general trends at the state level and any improvements in ESSPIN schools, depending on the scale of ESSPIN roll-out. The hope is that ESSPIN state-level interventions combined with the SIP will lead to an improvement in state-wide averages.
- Comparison of groups in CS2, according to whether they benefited from ESSPIN or did not. For Kwara this differentiation is made for teacher training, and for the Output stream 4 categories which target inclusiveness, SBMCs, school quality and pupil learning. Here we expect the results to be better for schools which have benefited from ESSPIN. If a school entered the programme in 2013/14, we count them as non-ESSPIN, as we would not expect the support to have impacted on the indicators yet.
- Comparison of schools which benefited from support in 2011/12 and 2012/13 with those which did not, to see whether they improved more or faster between 2012 and 2014. Here we expect the supported schools to improve relative to other schools over the two years. Again, in Kwara this comparison is made for teacher training, and the Output stream 4 categories which target inclusiveness, SBMCs, school quality and pupil learning.

3.1 Teacher competence

3.1.1 Main analysis

The ESSPIN logframe sets four criteria for judging competence of teachers (Box 4). A teacher who teaches English or maths is defined as competent if he or she meets at least three of these, while teachers of other subjects are exempted from one of the four criteria (knowledge of the English or maths curriculum) and defined as competent if they meet two of the remaining three criteria.

For CS2, a fifth criterion was added, based on teacher test results. Teachers are defined as competent if they are competent according to the original criteria, and can also score at least 50% in primary school-level literacy and numeracy tests.

Table 6 compares the results for Kwara's teachers in CS1 and CS2. (The fifth criterion is not available here as teacher tests were not conducted as part of CS1.) There was a significant improvement in the number of teachers using teaching aids in the lesson. By contrast, knowledge of the English or mathematics curriculum declined from 79% to 48%, although there is some possibility of measurement error in this indicator.² Overall in Kwara, the proportion of teachers meeting the (CS1) competence standard is now 74%, but this is not a significant change from CS1.

² CS2 introduced clearer guidance about which grade of the curriculum teachers should be asked about, 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.

Box 4. Logframe standard for teacher competence

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

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

For CS2, a new stricter indicator of teacher competence has been introduced. This excludes reading from or writing on, or having pupils copy from, the blackboard as a use of a teaching aid, and adds a fifth criterion:

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

We also calculate a continuous competence score, based on the number of criteria met by each teacher. A teacher who meets all of the three or four criteria would score 100%, while a teacher who meets none of them would score 0%. The competence score is also not significantly different in CS2 than in CS1. The absence of progress between CS1 and CS2 in Kwara may reflect the fact that the ESSPIN teacher training was in fact conducted in 2009/10 and 2010/11 (see Annex A), so any initial step change may have occurred before CS1 in 2012. Additional teacher training in 2013/14 is too recent to have had an effect by the time of CS2.

Table 6. Kwara: Teacher competence in CS1 and CS2

| | CS1 | CS2 | |
|-------------------------------------------|------|------|---|
| (1) Knowledge of English/maths curriculum | 79 | 48 | - |
| (2) Use of 1+ teaching aid | 89.4 | 98.4 | + |
| (3) Praise more than reprimand | 83.9 | 88 | |
| (4) Assigns 2+ ind./group task | 61.1 | 66.1 | |
| Competence score (CS1 version) | 77.4 | 75.2 | |
| Teacher competence standard (CS1) | 84.7 | 73.9 | |

Note: + = significant improvement between 2012 and 2014; - = significant worsening between 2012 and 2014 (using a t-test; p < .05)

It is part of ESSPIN's model that trained teachers should share their training with the rest of the school staff, with support from headteachers. Nevertheless, it is interesting to examine whether those teachers who personally received training from ESSPIN are now stronger than those who received the training through their colleagues. Comparing our indicators between these two groups (Table 7), we find no evidence to suggest that teachers perform better if they have directly received ESSPIN training as opposed to being in an ESSPIN-supported school. None of the results for the individual criteria, or the literacy and numeracy tests, are significantly different between teachers who individually received training and those who did not. Overall, 40% of trained teachers meet the new competence standard, which is not significantly different from the 36% of non-ESSPIN-trained teachers who meet the standard.

Table 7. Kwara: Teacher competence in CS2, by intervention group

| | (1) ESSPIN school | (2) ESSPIN-trained |
|--------------------------------------------------------------|-------------------|--------------------|
| Knowledge of English/maths curriculum | 57.3 | 39.1 |
| Use of 1+ teaching aid | 99.7 | 97.1 |
| Use of 1+ teaching aid excl. read/write/copy from blackboard | 80.5 | 90.3 |
| Praise more than reprimand | 79.6 | 96.4 |
| Assigns 2+ ind./group task | 69 | 63.1 |
| Literacy score (%) | 47.8 | 51.9 |
| Numeracy score (%) | 70.1 | 72.1 |
| Passes literacy and numeracy test | 47.6 | 54.8 |
| Competence score (CS1 version) | 76.8 | 73.6 |
| Teacher competence standard (CS1) | 74.9 | 72.9 |
| Competence score (CS2 version) | 70.5 | 71.4 |
| Teacher competence standard (CS2) | 35.9 | 40 |

Note: 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%; + / - indicate a significant difference between the results for teachers who individually received training and those who did not.

Did the teachers who were direct training recipients show more progress between 2012 and 2014 than the teachers who were in ESSPIN schools but who were not personally trained? Surprisingly, our results suggest the reverse (Table 8). Teacher competence scores improved more among teachers who were in ESSPIN schools but did not directly receive training themselves, and the difference is statistically significant.

Table 8. Kwara: Teacher competence difference in differences (comparison of means)

| Teacher competence scores (CS1 version) | (1) ESSPIN school | (2) ESSPIN-trained |
|-----------------------------------------|-------------------|--------------------|
| CS1 | 71.9 | 82.6 |
| CS2 | 76.8 | 73.6 |
| difference | 4.9 | -9* |

Note: * indicates a significantly different difference than those in ESSPIN schools but not trained ($p < .05$).

Overall in Kwara the findings suggest that there was no significant improvement in teacher competence between 2012 and 2014, and that teachers who have received training through ESSPIN are no more likely to meet the competence standard than those who were in ESSPIN schools but did not directly receive training. The untrained group appears to have improved between 2012 and 2014. This pattern may reflect the fact that direct training of teachers by ESSPIN ceased during 2011/12 and 2012/13, and the only interventions under ESSPIN's Output 3 were a large number of school visits. The improvement among teachers who did not receive training directly is consistent with the idea of trained teachers sharing resources so that their colleagues also benefit. However, further exploration is needed into why teachers who directly received training before 2012 did not continue to improve. This finding has implications for ESSPIN's sustainability. ESSPIN's school-level interventions are supposed to bring about continued improvement beyond the period of direct intervention. We do not have evidence for such continued improvement among trained teachers, who are not significantly better, by any of our measures, than the non-trained teachers. This could result, for example, from changes in the

motivation of trained teachers after they have ceased receiving training, or from the loss of the best trained teachers to other schools or professions.

3.1.2 Findings from the teacher content knowledge tests

The teacher tests included items pitched at primary school grades 1 to 5 and focusing on different areas: foundational skills for teaching literacy; writing; reading; grammar; number concepts; calculation; and other numeracy skills. In Kwara, on average, teachers scored more than double in reading and grammar than in writing (Figure 2). In mathematics, teachers performed better in number concepts than in calculation and on other numeracy items. As would be expected, the teachers' ability to answer the questions falls as the grade level of the questions increases (Figure 3). For grade 2 level knowledge and higher, the mathematics items were notably easier for teachers in Kwara than the English questions.

Figure 2. Kwara: Teacher test scores across domains of learning

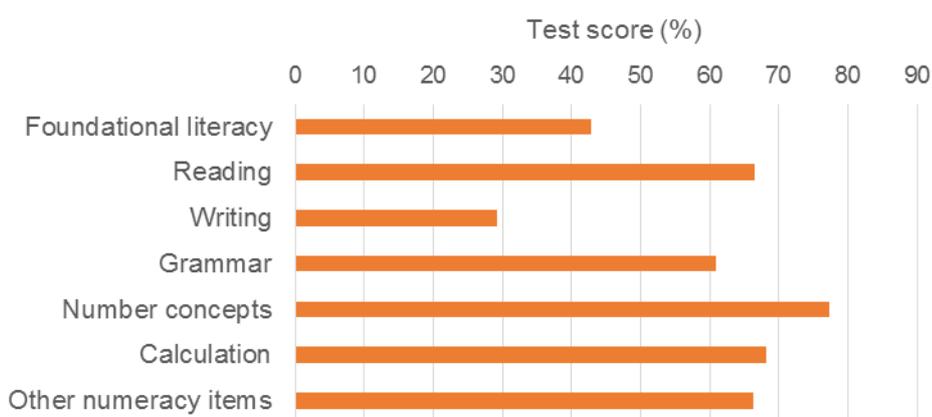
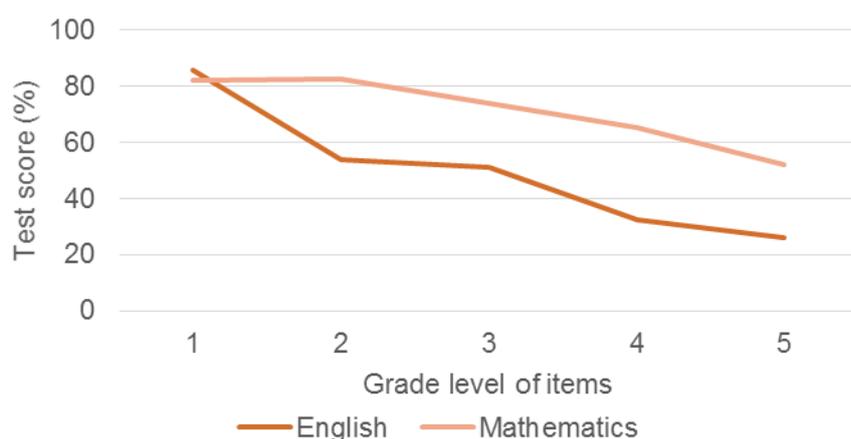


Figure 3. Kwara: Teacher test scores by grade



3.2 Headteacher effectiveness

The ESSPIN logframe defines headteacher effectiveness in terms of seven criteria (Box 5). These reflect both activities by the headteacher and behaviour across the teachers and pupils, such as agreement on what time the school should open (criterion 4), presence in class at the beginning of the school day (criterion 5), and appropriate break and lesson durations (criteria 6 and 7).

Box 5. Logframe standard for headteacher effectiveness

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

- 1) Carry out two or more lesson observations in the past two weeks;
- 2) Hold four or more professional development meetings since the start of the 2011/12 or 2013/14 school year (NB: the survey took place more than nine months into the school year);
- 3) School has a teacher attendance book and the headteacher recalls at least two actions taken to promote teacher attendance;
- 4) Clear school opening time: more than 50% of pupils sampled agree on the school opening time and more than 50% of teachers sampled agree on the school opening time;
- 5) More than 50% of classes are in their classroom with their teacher within 30 minutes of school opening time;
- 6) Length of morning break is 35 minutes or less, except in Enugu when it must be 15 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).

Overall, in Kwara there was very little change in the proportion of headteachers who met the criteria for effectiveness between CS1 and CS2 (Table 9). Out of the seven criteria, one improved (conducting professional development meetings) but one worsened (taking action to address teacher attendance) over the period. In CS2, 22% of headteachers met the overall standard, which is not statistically significantly different from the CS1 finding (19%). Again, as with the teacher results, this may reflect the fact that the leadership training was conducted in 2009/10 and 2010/11, suggesting that any improvement that was due to the training may have occurred before the first Composite Survey, making no further impact between CS1 and CS2.

Table 9. Kwara: Headteacher effectiveness in CS1 and CS2

| | CS1 | CS2 | |
|------------------------------------------|------|------|---|
| (1) Lesson observations | 24.6 | 17.9 | |
| (2) Professional development meetings | 20 | 37.6 | + |
| (3) Action on teacher attendance | 73.3 | 45.7 | - |
| (4) Clear opening time | 43 | 36.4 | |
| (5) In class on time | 89.2 | 74.1 | |
| (6) Appropriate morning break | 88 | 78.8 | |
| (7) Appropriate lesson length | 15 | 27.3 | |
| Number of criteria fulfilled (/7) | 3.5 | 3.2 | |
| Effective headteacher (5/7 criteria met) | 18.8 | 22.4 | |

Note: + = significant improvement between 2012 and 2014; - = significant worsening between 2012 and 2014 (using a t-test; $p < .05$).

3.3 School development planning

The definition of effective school development planning depends on five criteria (Box 6). On all five criteria, Kwara improved between 2012 and 2014, with the difference reaching statistical significance for two of the five indicators: having a SDP, and whether the plan contains activities to strengthen teaching and learning (Table 10). The average number of criteria fulfilled doubled from 1.1 in 2012 to 2.2 in 2014, and the number of schools meeting the standard increased from 6% to

25%. A large majority (80%) of schools now have an SDP of some form, and 21% could present evidence of having carried out activities in it, compared to 11% in 2012.

Box 6. Logframe standard for effective 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).

Table 10. Kwara: SDP effectiveness in CS1 and CS2

| | CS1 | CS2 | |
|--------------------------------------------------------------------|------|------|---|
| (1) Written evidence of school self-evaluation process | 29.8 | 47.2 | |
| (2) SDP available | 43.5 | 80 | + |
| (3) SDP contains 3+ activities to strengthen teaching and learning | 24.6 | 46.3 | + |
| (4) Evidence that 4+ activities from SDP carried out | 11.2 | 21.3 | |
| (5) Cashbook up-to-date | 13.8 | 25.4 | |
| Number of SDP criteria fulfilled (/5) | 1.1 | 2.2 | + |
| School meets effective school development planning standard | 6.4 | 25.0 | + |

3.4 School inclusiveness: meeting the needs of all pupils

The school inclusiveness standard depends on meeting three out of four criteria (Box 7), and schools are defined as partially meeting the standard if two criteria are met. In Kwara a number of the inclusiveness criteria indicators actually declined between 2012 and 2014, although there was an increase in schools with activities to improve access in their SDP (Table 11). The reductions led to the overall proportion of schools meeting the standard falling significantly from 45% to 20%.

Box 7. Standard for school inclusiveness (meeting the needs of all pupils)

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) Headteacher 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 between 60% and 40% of total engagements meet the criterion).

Table 11. Kwara: School inclusiveness in CS1 and CS2

| | CS1 | CS2 | |
|------------------------------------------------------------------------|------|------|---|
| (1) 3+ actions to improve attendance | 76.4 | 30.3 | - |
| (2) 2+ activities in SDP to improve access for disadvantaged children | 14.7 | 47.8 | + |
| (3) >50% of teachers use 2+ assessment methods | 78.3 | 91.8 | |
| (4) >50% of teachers spatially inclusive and >50% are gender inclusive | 52.1 | 17.9 | - |
| Number of inclusiveness criteria fulfilled (/4) | 2.2 | 1.9 | |
| Inclusiveness score | 79.3 | 69.7 | - |
| School partially met inclusiveness standard (2–4 criteria out of 4) | 78.4 | 67.3 | |
| School fully met inclusiveness standard (3–4 criteria out of 4) | 44.7 | 20.1 | - |

Note: + = significant improvement between 2012 and 2014; - = significant worsening between 2012 and 2014 (using a t-test; $p < .05$).

Focusing on CS2 schools, we separate the schools according to the level of support they received under Output stream 4 of ESSPIN. More schools with Output 4 intervention had carried out actions to address pupil attendance (65%) than schools with little or no Output 4 intervention (24%) (Table 12). Significantly more of the Output 4 schools met the inclusiveness standard (46%) than other schools (16%).

Difference in difference analysis of schools in Kwara depending on the level of ESSPIN Output 4 intervention has been carried out by comparing the change in inclusiveness score between the schools which had full Output 4 intervention and those that did not (Table 13). The negative change in the mean for both groups of schools tells us that both sets of schools became less inclusive between 2012 and 2014. However, the direction of the results suggests that schools which received more ESSPIN Output stream 4 support worsened less than other schools. This is shown by the less negative change in Table 13 (although the difference was not found to be statistically significant). In other words, the ESSPIN support may have helped counteract a drop in inclusiveness. Similar results are found when using a regression method for difference in difference analysis (Annex C).

Table 12. Kwara: School inclusiveness in CS2, by Output stream 4 intervention gro

| | (1) No / little intervention | (2) Full intervention | |
|-------------------------------------------------------------------------|------------------------------|-----------------------|---|
| <i>Inclusiveness criteria</i> | | | |
| (1) 3+ actions to improve attendance | 24 | 64.5 | + |
| (2) 2+ activities in SDP to improve access for disadvantaged children | 47.4 | 50.4 | |
| (3) >50% of teachers use 2+ assessment methods | 90.3 | 100 | |
| (4) >50% of teachers spatially inclusive and >50% are gender inclusive | 15.1 | 32.6 | |
| <i>Overall inclusiveness standard</i> | | | |
| Number of inclusiveness criteria fulfilled (/4) | 1.8 | 2.5 | + |
| Inclusiveness score | 69.6 | 69.8 | |
| School partially met inclusiveness standard (2–4 criteria out of 4) | 62.2 | 94.6 | + |
| School fully met inclusiveness standard (3–4 criteria out of 4) | 15.5 | 45.5 | + |
| <i>Detailed</i> | | | |
| Number of actions to improve attendance | 2.2 | 2.8 | + |
| Number of activities on access for disadvantaged children | 1.7 | 1.8 | |
| Average number of assessment methods used | 2.3 | 2 | |
| Average number of zones participating in lessons | 3.1 | 3.5 | |
| Average gender equity score (0=completely unequal, 100=perfectly equal) | 88.2 | 81.6 | |

Note: + indicates a significant positive difference between non-ESSPIN and ESSPIN schools. The gender equity score for a teacher is $100 - 100 \times \text{abs}(\frac{g}{g+b} - \frac{G}{G+B})$ where g is the number of girls who participate, b is the number of boys who participate, G is the number of girls present in the class, and B is the number of boys present in the class. It is expressed as a percentage score. For a lesson where the proportion of girls and boys participating is exactly equal to the proportion of girls and boys sitting in the lesson, the gender equity score will be 100; for a lesson where no boys participate, or no girls participate, the score will be zero.

Table 13. Kwara: School inclusiveness difference in differences (comparison of means)

| Inclusiveness score | (i) No/little intervention | (ii) Full intervention |
|---------------------|----------------------------|------------------------|
| CS1 | 80 | 76 |
| CS2 | 69.6 | 69.8 |
| Difference | -10.3 | -6.2 |

Note: * indicates a significant difference in differences compared to the schools with no or little Output 4 intervention ($p < .05$).

3.5 SBMC functionality and inclusiveness

ESSPIN conducted qualitative research into SBMCs and community engagement in education in five ESSPIN states in 2009 (ESSPIN 2009) (Jigawa, Kaduna, Kano, Kwara and Lagos). This research suggested that SBMCs were not functioning well: there was a lack of clarity and understanding over the SBMC's role and responsibilities; they lacked the financial resources to support schools in the ways that Local Government Education Areas (LGEAs) often expected them

to; community members were sometimes excluded by local elites; and there was little participation by women and children, despite guidelines requiring their inclusion.

In this context, SBMCs were starting from a low base and with substantial sociocultural barriers to be overcome to reach functionality and inclusive participation. ESSPIN has aimed to improve community involvement in schools through functioning SBMCs and increased women's and children's participation, with a number of interventions under its Output stream 4 (see Annex B).

In Kwara, almost all of the schools sampled did have a committee in place. Only two of the CS1 sample did not have an SBMC; all schools had an SBMC in CS2 (Table 14). This does not mean that all the SBMCs are functional or inclusive, however. The following sections use criteria and standards defined by the ESSPIN logframe to examine SBMC functionality and the extent to which SBMCs are inclusive of women and children.

Table 14. Kwara: Sample size of schools with SBMCs

| | CS1 | CS2 |
|-------------------------------------|-----|-----|
| Schools sampled in Kwara | 102 | 105 |
| Schools with SBMCs sampled in Kwara | 100 | 105 |

3.5.1 SBMC functionality

There are nine criteria used to assess SBMC functionality, of which five must be fulfilled to meet the logframe standard (Box 8). In Kwara there was a significant improvement in seven of the nine criteria between CS1 and CS2 (Table 15). Of these the largest improvements were in SBMC networking (criterion 4) and those which have a women's committee and a children's committee (criteria 6 and 7) – more than double the number of schools met these criteria in 2014 than in 2012. Overall the number of schools meeting the SBMC functionality standard more than doubled from 30% to 75%, in just two years.

Two additional inclusiveness-related criteria not included in the CS1 report are also examined in this section: whether the SBMC did anything to support commonly excluded groups, and whether it raised issues of children's exclusion from school with the community, LGEA or state government. The first of these fell but the second improved over the period in Kwara.

Box 8. Logframe standard 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:³

- 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 LGEAs on education service delivery issues (written or physical evidence);
- 6) An SBMC women's committee exists (written or physical evidence);
- 7) An SBMC children's committee exists (written or physical evidence);
- 8) SBMC contributed resources for the school (written or physical evidence); and
- 9) The SBMC chair visited the school at least three times from the start of the current school year (written evidence).

Table 15. Kwara: SBMC functionality in CS1 and CS2

| | CS1 | CS2 | |
|--------------------------------------------------|------|------|---|
| (1) 2+ meetings this school year | 41.3 | 71.7 | + |
| (2) Conducted awareness raising | 56.7 | 86.8 | + |
| (3) Addressed exclusion | 45.8 | 67.8 | + |
| (4) Networked with CBOs/institutions/other SBMCs | 18.5 | 90.8 | + |
| (5) Interacted with LGEA | 23.1 | 35.4 | |
| (6) Has women's committee | 27 | 68.1 | + |
| (7) Has children's committee | 14.6 | 57.7 | + |
| (8) Contributed resources for school | 48.8 | 71.4 | + |
| (9) Chair visited school 3+ times | 27.2 | 35.4 | |
| Schools meeting functioning SBMC standard | 29.6 | 75.1 | + |
| Number of SBMC functionality criteria met (/9) | 3 | 5.9 | + |
| <i>Additional criteria</i> | | | |
| Action for commonly excluded groups | 37.5 | 16.4 | - |
| Raised issue of children's exclusion | 4.9 | 13.3 | + |

Note: + = significant improvement between 2012 and 2014; - = significant worsening between 2012 and 2014 (using a t-test; $p < .05$).

In 2014, Kwara's schools which received more ESSPIN Output 4 intervention performed more effectively on all the SBMC functionality criteria than those that received little or no Output 4 intervention, and seven out of nine of them significantly so (Table 16). Notably, 100% of the Output 4 schools met the overall standard, compared with 70% of the schools that were not in the initial phase of Output 4 roll-out. Schools with Output 4 interventions also had better results than the comparison schools in terms of our inclusiveness criteria: they were much more likely to have taken action for commonly excluded groups, or to have raised issues of children's exclusion.

³ A slightly different standard with 10 criteria was used in CS1. The new standard with nine criteria was applied to both the CS1 and CS2 data.

Table 16. Kwara: SBMC functionality in CS2, by Output stream 4 intervention group

| | (i) No/little Output 4 intervention | (ii) Full intervention | |
|------------------------------------------------|-------------------------------------|------------------------|---|
| (1) 2+ meetings this school year | 68.8 | 87.4 | |
| (2) Conducted awareness raising | 84.8 | 98.3 | + |
| (3) Addressed exclusion | 63.4 | 91.0 | + |
| (4) Networked | 89.3 | 98.6 | |
| (5) Interacted with LGEA | 30.8 | 60.2 | + |
| (6) Has women's committee | 62.1 | 100 | + |
| (7) Has children's committee | 50.7 | 94.8 | + |
| (8) Contributed resources for school | 67.1 | 94.2 | + |
| (9) Chair visited school 3+ times | 28.2 | 73.2 | + |
| Standard G: functioning SBMC | 70.4 | 100 | + |
| Number of SBMC functionality criteria met (/9) | 5.5 | 8.0 | + |
| <i>Additional criteria</i> | | | |
| Action for commonly excluded groups | 8.3 | 60.0 | + |
| Raised issue of children's exclusion | 8.4 | 39.8 | + |

Note: + indicates a significant positive difference between schools with full output 4 intervention and those with little or no output 4 intervention.

We are interested to know whether the schools which received more support under ESSPIN Output stream 4 improved faster than schools with less support between 2012 and 2014. Both groups improved over the period, meeting at least two additional criteria. There was no statistically significant difference between Output 4 and comparison schools in the amount of improvement (Table 17). A regression analysis, shown in Annex C, reinforces this finding. However, this may be partly a result of our scale score, which gives equal weight to each of the nine criteria. Some of the Output 4 schools were already meeting most of the criteria in 2012, leaving little room for further improvement between 2012 and 2014. The schools that did not receive Output 4 intervention started from a lower base, and so had much greater scope for measurable improvement.

Table 17. Kwara: SBMC functionality difference in differences (comparison of means)

| Number criteria met (out of 9) | (i) No/little Output 4 intervention | (ii) Full intervention |
|--------------------------------|-------------------------------------|------------------------|
| CS1 | 2.4 | 5.9 |
| CS2 | 5.5 | 8.0 |
| difference | 3.0 | 2.0 |

Note: * indicates a significant difference in differences compared to the schools with no or little output 4 intervention ($p < .05$).

3.5.2 Women's inclusiveness

The 2009 study of SBMCs found that community members were excluded from the process by local elites, and as such SBMCs were little known about and lacked a link to the community. In addition, the requirement (as stated in the guidelines) for participation by women and students was often ignored where this was felt inappropriate in the local culture. In this section and the following section, we examine the extent to which SBMCs were inclusive of women's and children's concerns in 2014. We measure SBMC women's inclusiveness using four criteria (Box 6).

In Kwara there was a significant improvement between CS1 and CS2 on all of the criteria, such as the number who had a women's committee meeting take place, which increased from 5% in 2012 to 74% in 2014. The number of SBMCs meeting the standard rose from 27% to 71%, a significant improvement in women's inclusiveness (Table 18).

Box 9. Logframe standard for SBMC women's inclusiveness

| |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p>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:</p> <ol style="list-style-type: none"> 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.⁴ |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

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

| | CS1 | CS2 | |
|-----------------------------------------------------|------|------|---|
| (1) At least one woman attended 2+ meetings (%) | 37.1 | 58.5 | + |
| (2) Female member raised an issue (%) | 44.2 | 84.8 | + |
| (3) Issue raised by female member led to action (%) | 44.9 | 65.2 | + |
| (4) Women's committee met (%) | 5.2 | 73.6 | + |
| Number of criteria met | 1.3 | 2.8 | + |
| Meets standard (3/4 criteria) | 26.6 | 70.8 | + |

Note: Schools that did not have SBMCs at all in CS1 are excluded from the analysis; + = significant improvement between 2012 and 2014; - = significant worsening between 2012 and 2014 (using a t-test; $p < .05$).

Women's inclusiveness was found to be higher and more significant in schools which had received more Output stream 4 interventions than in those that received fewer, for all the criteria in CS2 (Table 19). Overall, 93% of Kwara's schools which received more ESSPIN Output stream 4 support met the standard; 67% of the comparison schools did.

Table 19. Kwara: SBMC women's inclusiveness in CS2, by Output stream 4 intervention group

| | (i) No/little Output 4 intervention | (ii) Full intervention | |
|-----------------------------------------------------|-------------------------------------|------------------------|---|
| (1) At least one woman attended 2+ meetings (%) | 54 | 82.8 | + |
| (2) Female member raised an issue (%) | 82 | 100 | + |
| (3) Issue raised by female member led to action (%) | 61.4 | 85.8 | + |
| (4) Women's committee met (%) | 69 | 98.3 | + |
| Number of criteria met | 2.7 | 3.7 | + |
| Meets standard (3/4 criteria) | 66.8 | 92.6 | + |

Note: + indicates a significant positive difference between schools with full output 4 intervention and those with little or no output 4 intervention.

⁴ This criterion has been slightly altered since CS1, where it also required the women's committee to have a female leader.

The comparison of means method of difference in differences (Table 20) actually finds that although all schools improved in women's inclusiveness over the two years, the schools receiving less Output 4 intervention actually improved more than those with full intervention. A regression analysis supports this conclusion (see Annex C). As with SBMC functionality above, it may reflect the fact that the treatment schools were already inclusive of women in 2012, with 2.8 out of the four criteria met, giving them less room for improvement, while the no intervention schools had further to catch up.

Table 20. Kwara: SBMC women's inclusiveness difference in differences (comparison of means)

| Number criteria met (out of 4) | (i) No/little Output 4 intervention | (ii) Full intervention |
|--------------------------------|-------------------------------------|------------------------|
| CS1 | 0.9 | 2.8 |
| CS2 | 2.7 | 3.7 |
| Difference | 1.7 | 0.9* |

Note: Schools that did not have SBMCs at all in CS1 are excluded from the analysis; * indicates a significant difference in differences compared to the schools with no or little output 4 intervention ($p < .05$).

3.5.3 Children's inclusiveness

Earlier qualitative research (ESSPIN, 2009) found that many SBMCs did not allow the participation of children, and that where they had student members, they were not always able to be invited or may not have been comfortable voicing opinions in meetings. In this section, we examine whether SBMCs have improved in the extent to which they are inclusive of children, in accordance with guidelines on how they are supposed to operate. There are four criteria in the standard on SBMC children's inclusiveness.

In Kwara two of the four criteria saw significant improvements between CS1 and CS2 – there were more SBMCs in which a child had raised an issue, and more children's committees had met (Table 21). More than twice the number of SBMCs met the children's inclusiveness standard in 2014 (27%) compared with 2012 (11%).

Box 10. Logframe standard for SBMC children's inclusiveness

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.⁵

Table 21. Kwara: SBMC children's inclusiveness in CS1 and CS2

| | CS1 | CS2 | |
|----------------------------------------------------------------|------|------|---|
| (1) Child attended 2+ meetings (%) | 25.1 | 24.5 | |
| (2) Child raised an issue (%) | 27 | 56.3 | + |
| (3) Issue raised by child led to action (%) | 20.7 | 31.5 | |
| (4) Children's committee met and has a trained facilitator (%) | 5.5 | 34.3 | + |
| Number of criteria met | 0.7 | 1.5 | + |
| Meets standard (3/4 criteria) (%) | 11.1 | 26.6 | + |

Note: Schools that did not have SBMCs at all in CS1 are excluded from the analysis; + = significant improvement between 2012 and 2014; - = significant worsening between 2012 and 2014 (using a t-test; $p < .05$)

Looking at children's inclusiveness in CS2, the schools receiving maximum ESSPIN Output stream 4 intervention had higher performance than those with little or no intervention, on all the indicators – as would be expected (Table 22). This difference was significant in terms of children attending SBMC meetings, raising an issue in the SBMC and the children's committee holding meetings. Overall, 64% of full-intervention schools met the standard for SBMC children's inclusiveness while 20% of the comparison schools did.

Table 22. Kwara: SBMC children's inclusiveness in CS2, by Output stream 4 intervention group

| | (i) No/little Output 4 intervention | (ii) Full intervention | |
|----------------------------------------------------------------|-------------------------------------|------------------------|---|
| (1) Child attended 2+ meetings (%) | 18.6 | 56.4 | + |
| (2) Child raised an issue (%) | 50.3 | 88 | + |
| (3) Issue raised by child led to action (%) | 27.6 | 51.6 | |
| (4) Children's committee met and has a trained facilitator (%) | 27.1 | 73.5 | + |
| Number of criteria met | 1.2 | 2.7 | + |
| Meets standard (3/4 criteria) (%) | 19.6 | 64.4 | + |

Note: + indicates a significant positive difference between schools with full Output 4 intervention and those with little or no Output 4 intervention.

The difference in differences analysis finds evidence that while all schools improved, the schools with more Output 4 intervention improved at the same rate between 2012 and 2014 as the

⁵ 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.

comparison schools (Table 23). Again the comparison schools had much more room to improve than those with maximum Output stream 4 intervention.

Table 23. Kwara: SBMC children's inclusiveness difference in differences (comparison of means)

| Number of criteria fulfilled (out of 4) | (i) No/little Output 4 intervention | (ii) Full intervention |
|-----------------------------------------|-------------------------------------|------------------------|
| CS1 | 0.4 | 2.2 |
| CS2 | 1.2 | 2.7 |
| Difference | 0.8 | 0.5 |

Note: Schools that did not have SBMCs at all in CS1 are excluded from the analysis; * indicates a significant difference in differences compared to the schools with no or little Output 4 intervention ($p < .05$).

Overall, Kwara saw substantial improvements in the inclusiveness of its SBMCs between 2012 and 2014. The results show that schools that received more support under Output stream 4 of ESSPIN did have more inclusive SBMCs. On some of our indicators, schools receiving the most support are reaching the top of our scale, so there is little scope for them to improve further in ways that can be measured with the existing sets of criteria.

3.6 School quality

Overall school quality is measured as a combination of the standards on teacher competence, headteacher effectiveness, school development planning, and SBMC functionality. A high-quality school is defined as one that meets the teacher competence standard and at least two of the other standards (Box 11).

The results for Kwara show that there has been a significant improvement in school quality between CS1 and CS2 (Table 24). The proportion of schools meeting the standard increased from 11% to 29%. We also use a 'quality score' indicator, which is an average of the continuous indicators developed in the previous sections for teacher competence, headteacher effectiveness, school development planning, and SBMC functionality. There was a notable increase in this quality score from 46% in CS1 to 57% in CS2.

Box 11. 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 50% of sampled teachers are competent);
- 2) Headteacher effectiveness standard;
- 3) School development planning effectiveness standard; and
- 4) SBMC functionality standard.

As the teacher competence standard has changed between CS1 and CS2 – with teachers required to score a minimum of 50% in both English and mathematics tests to pass the competence standards – we report both 'CS1' and stricter 'CS2' versions of the overall quality standard.

Table 24. Kwara: School quality in CS1 and CS2

| | CS1 | CS2 | |
|---------------------------------------------|------|------|---|
| Meets three or four standards (CS1 version) | 11.4 | 28.9 | + |
| Quality score (CS1 version) | 45.9 | 56.7 | + |

Note: + = significant improvement between 2012 and 2014; - = significant worsening between 2012 and 2014 (using a t-test; $p < .05$).

Whether a school had received no or minimal Output stream 4 intervention or had received the maximum, there was little difference in the proportion of schools meeting the school quality standard (Table 25). This may be expected given that all schools received the same Output stream 3 support (which target three of the four standards) and it is only the functioning SBMC standard which would be affected by varied Output stream 4 intervention. However, using our continuous indicator of school quality, scores are significantly higher (around 65%) in schools with more intervention than in those with little or no Output 4 intervention (54%).

Table 25. Kwara: School quality in CS2, by Output stream 4 intervention group

| | (i) Non-ESSPIN | (ii) ESSPIN | |
|---------------------------------------------|----------------|-------------|---|
| Meets three or four standards (CS1 version) | 28.3 | 32.9 | |
| Meets three or four standards (CS2 version) | 17.8 | 25 | |
| Quality score (CS1 version) | 55.3 | 65.2 | + |
| Quality score (CS2 version) | 53.9 | 65 | + |

Note: The CS2 version of the quality score and school quality standard reflect the 'strict' version of the teacher competence standard, where teachers are required to pass literacy and numeracy tests as well as fulfilling other criteria.

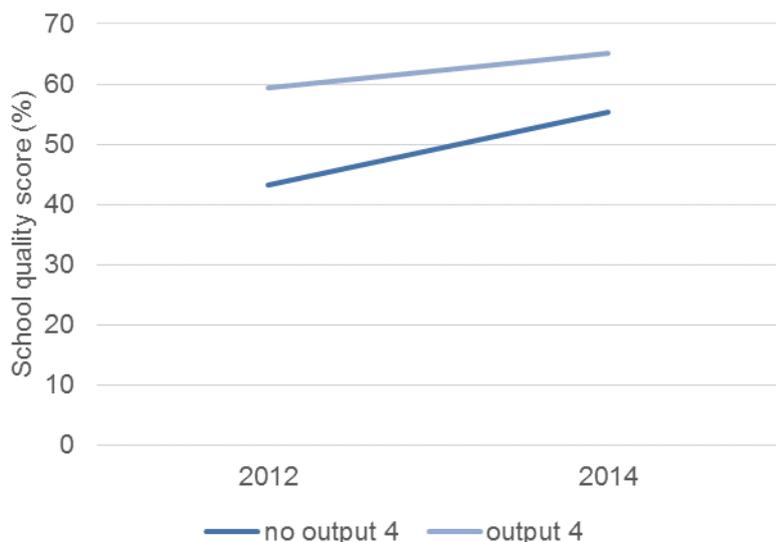
The comparison of changes in means shows that while all schools improved their quality score between 2012 and 2014, there is no evidence that schools with more Output stream 4 support improved any faster than other schools (Table 26). These results can be seen visually in Figure 4, with all schools improving by a similar amount but the Output stream 4 schools being at a higher level. Difference in differences was also conducted using a regression method and finds similar results (Annex C).

Table 26. Kwara: School quality difference in differences (comparison of means)

| School quality score | (i) No/little Output 4 intervention | (ii) Full intervention |
|----------------------|-------------------------------------|------------------------|
| CS1 | 43.2 | 59.3 |
| CS2 | 55.3 | 65.2 |
| difference | 12.1 | 5.9 |

Note: * indicates a significant difference in differences compared to the schools with no or little Output 4 intervention ($p < .05$).

Figure 4. Kwara: School quality in 2012 and 2014, by level of Output stream 4 intervention



3.7 Pupil learning achievement in English literacy and numeracy

The learning achievement indicators are included as indicators of the impact of the ESSPIN programme. Pupils were tested in grades 2 and 4, in literacy and numeracy. This section begins by following the same analysis conducted for other indicators: looking at the change in the state average between CS1 and CS2, the difference between schools with more or less Output 4 intervention in 2014, and the change in results between 2012 and 2014 for pupils in schools expected to improve more in this time compared with those expected to improve less. It then moves on to a more detailed look at the breakdown of pupil results. Here we look at how pupil scores were distributed in the 2014 tests, split between schools which received Output stream 4 support and those which did not. We also look at the average test scores on sub-scales of the tests, such as grasp-of-number concepts or addition and subtraction, and how this varies over time for the state average and for pupils from different types of schools.

3.7.1 Main analysis

From CS1 to CS2 the average pupil test scores in Kwara improved significantly in literacy, by over six percentage points in both grades 2 and 4 (Table 27). However, the scores for numeracy in grade 2 actually worsened, from 61% down to 52%.

As well as changes in average test scores, Table 27 shows the proportion of schools that met ESSPIN's logframe indicator. This is a relatively narrow indicator based on whether each child correctly answered a subset of questions in each test or not (see Cameron, 2015, Annex A.1). Very low numbers of pupils met most of these logframe indicators.

Table 27. Kwara: Test scores and proportion of children reaching logframe indicator in CS1 and CS2

| | Test | CS1 | CS2 | |
|------------------------|------|------|------|---|
| Test score (%) | L2 | 37.1 | 43.1 | + |
| | L4 | 36 | 42.7 | + |
| | N2 | 60.8 | 51.5 | - |
| | N4 | 41.7 | 39 | |
| Logframe indicator (%) | L2 | 1.9 | 0.4 | |
| | L4 | 1.1 | 4.1 | |
| | N2 | 20.6 | 4.1 | - |
| | N4 | 4.7 | 3.4 | |

Note: + = significant improvement between 2012 and 2014; - = significant worsening between 2012 and 2014 (using a t-test; $p < .05$).

Focusing on the CS2 data, there is little difference between the pupil test score results when separating schools by the amount of Output stream 4 intervention they received (Table 28). The average test scores are very similar in all subjects across the two groups, and the proportion of pupils meeting the logframe indicators are also similarly low for both groups.

ESSPIN does not argue that there is a direct causal link between inclusion and community participation (Output stream 4 interventions) and learning outcomes. Nevertheless, we have tested this to see if there is any association between Output 4 intervention and test scores, but find no evidence for it. It may still be the case that there is an indirect link that is not immediate enough to have had effects on the learning outcomes within the timeframe of the programme. As noted in section 1.2 above, there were rapid increases in enrolment and in pupil–teacher ratios in Kwara during 2009–2013, and these increases were largest in the schools that received most Output 4 intervention. This may have offset the effect on learning outcomes of any improvement in school functioning.

Table 28. Kwara: Test scores and proportion of children reaching logframe indicator in CS2, by Output stream 4 intervention group

| | Test | (i) No/little Output 4 intervention | (ii) Full intervention | |
|------------------------|------|-------------------------------------|------------------------|--|
| Test score (%) | L2 | 43.5 | 40.9 | |
| | L4 | 42.7 | 42.8 | |
| | N2 | 51.9 | 49.1 | |
| | N4 | 38.5 | 41.6 | |
| Logframe indicator (%) | L2 | 0.3 | 1 | |
| | L4 | 4.7 | 1.5 | |
| | N2 | 4.4 | 2.9 | |
| | N4 | 3.5 | 2.8 | |

Note: + indicates a significant positive difference between schools with full output 4 intervention and those with little or no Output 4 intervention.

As all schools in Kwara received the same support under ESSPIN for teacher and headteacher training (Output stream 3), we cannot estimate whether these activities improved pupil learning since we do not have a control group for comparison. However, we can look at the schools again according to the level of Output stream 4 intervention they received, to see whether these ESSPIN activities had an impact on change in pupil test scores between 2012 and 2014. Comparing

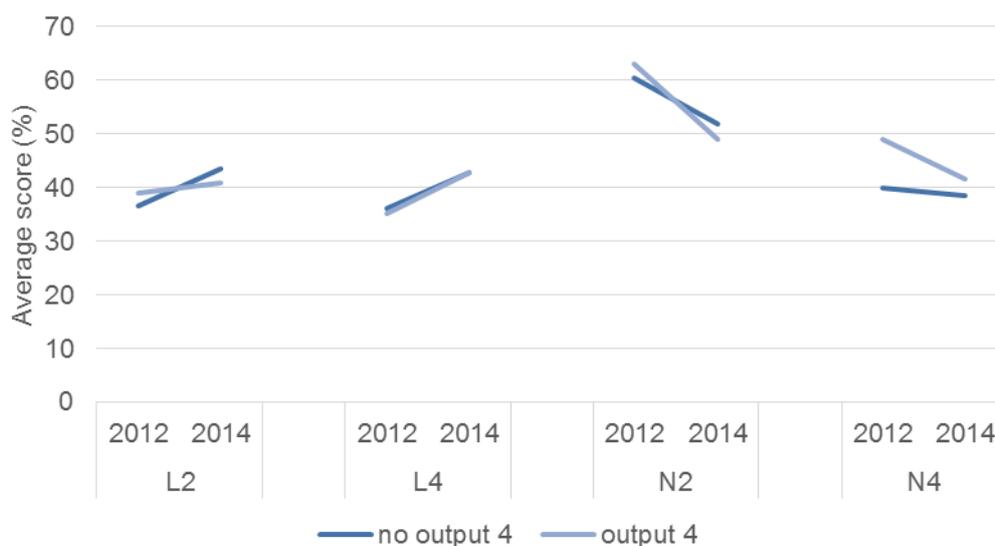
difference in differences (Table 29) confirms again the positive improvement in literacy and some worsening in numeracy. However, all the schools' pupil test scores changed at roughly the same rate for most subjects, giving no evidence that the Output stream 4 interventions have an additional impact on pupil test scores (this can be seen visually in Figure 5). There is some suggestion that scores in grade 2 numeracy improved more slowly (or in fact worsened faster) in schools which had benefited from more intervention over the period (this is confirmed in the regression method, shown in Annex C). Further investigation is needed to understand whether there learning outcomes are genuinely improving faster in schools that received less intervention. If so, it may be explained by factors such as enrolment increasing faster in schools that are receiving more Output stream 4 intervention.

Table 29. Kwara: Pupil test score difference in differences (comparison of means)

| Pupil test score | | (i) No/little Output 4 intervention | (ii) Full intervention |
|------------------|------------|-------------------------------------|------------------------|
| L2 | CS1 | 36.7 | 38.9 |
| | CS2 | 43.5 | 40.9 |
| | Difference | 6.8 | 2 |
| L4 | CS1 | 36.1 | 35.2 |
| | CS2 | 42.7 | 42.8 |
| | Difference | 6.5 | 7.5 |
| N2 | CS1 | 60.4 | 63 |
| | CS2 | 51.9 | 49.1 |
| | Difference | -8.5 | -13.9 |
| N4 | CS1 | 39.9 | 48.9 |
| | CS2 | 38.5 | 41.6 |
| | Difference | -1.5 | -7.2 |

Note: * indicates a significant difference in differences compared to the schools with no or little Output 4 intervention (p < .05).

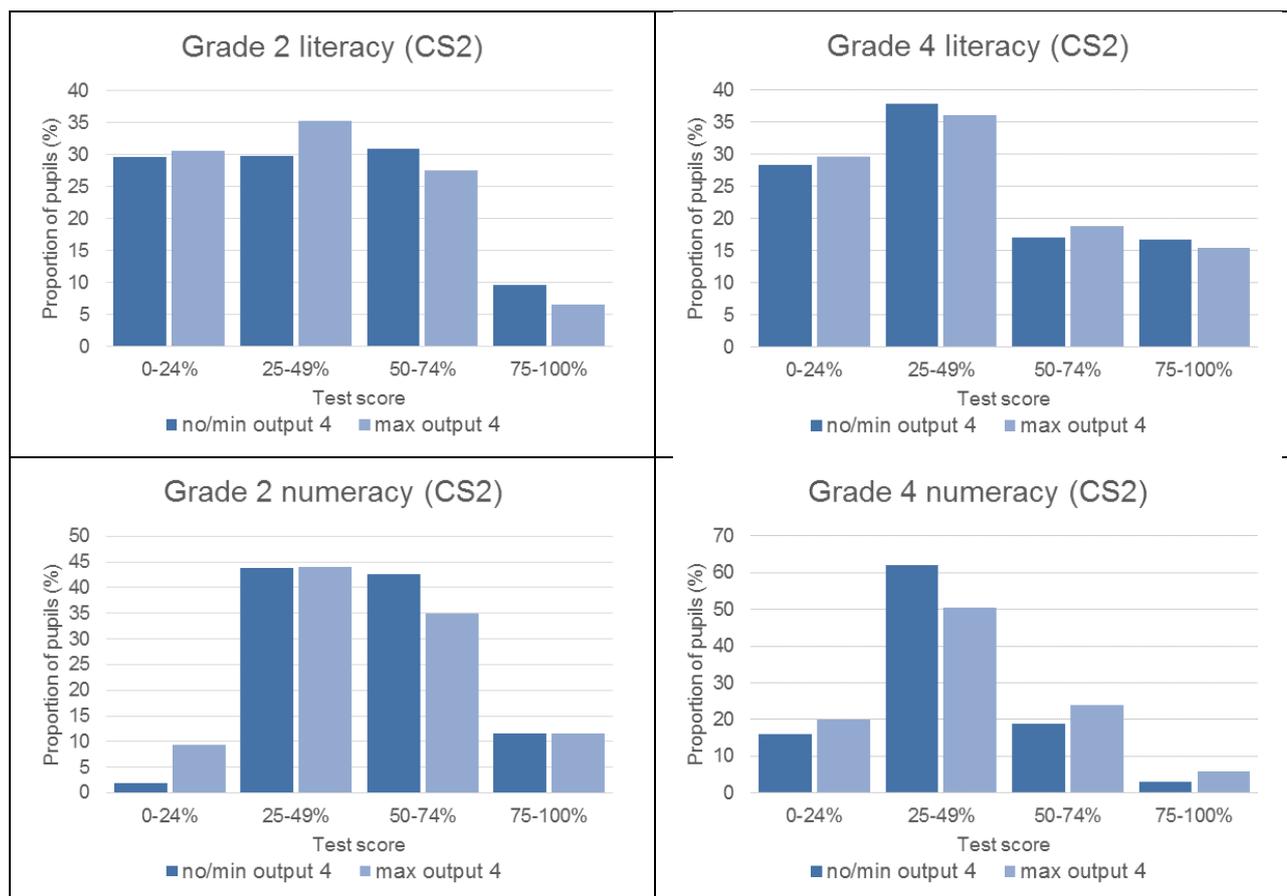
Figure 5. Kwara: Pupil test scores in schools by level of Output stream 4 intervention, 2012 and 2014



3.7.2 Distribution of test scores and sub-scale scores

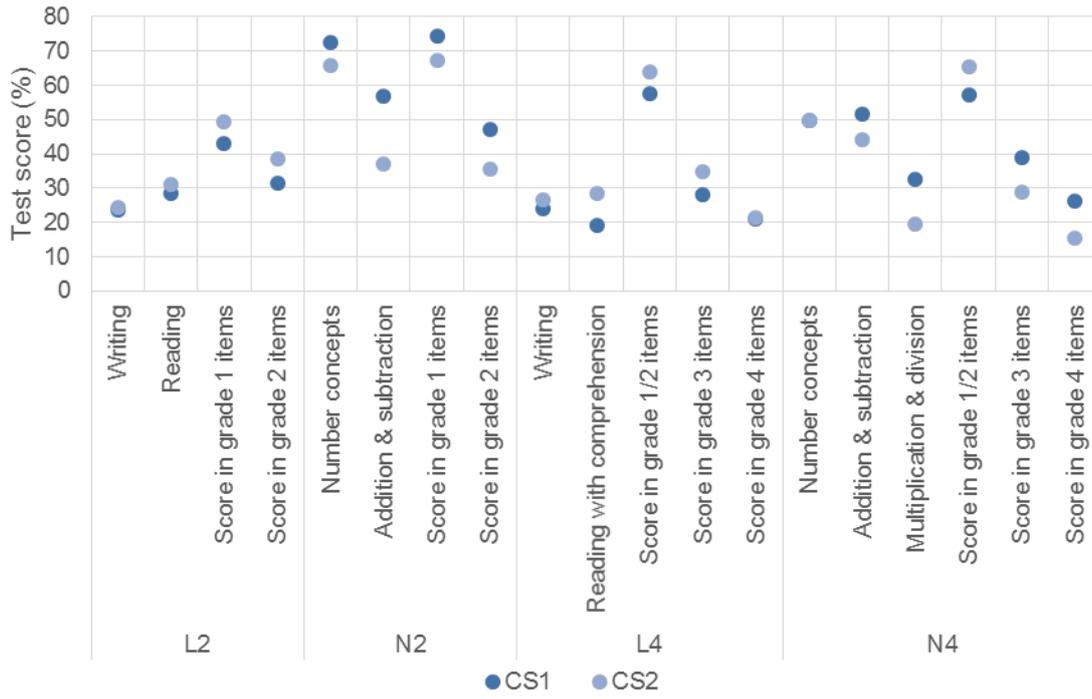
Looking at the distribution of pupils' test scores in 2014, there is no consistent pattern across the four tests in terms of difference between the two groups of schools or distribution (Figure 6). For literacy, particularly at grade 2, the results are spread fairly uniformly between the three lower bands, with a smaller proportion scoring in the 75–100% band. The numeracy results have far more pupils in the middle ranges, particularly the 25–49% band. There is little difference in the results when split by pupils from schools which received the maximum ESSPIN Output stream 4 intervention and those with no or minimal Output stream 4 intervention.

Figure 6. Kwara: Distribution of pupil test scores



Between CS1 and CS2 there was generally an improvement across the literacy sub-scale items in both grades 2 and 4 (Figure 7). For example, compared with 2012, in 2014 grade 2 pupils were better at answering the grade 1 and grade 2 level questions, and grade 4 pupils were better at reading with comprehension grade 1/2 and grade 3 level questions. However, across numeracy there were some drops in scores. In 2014, grade 2 pupils were less likely to answer number concepts, addition and subtraction and grade 1 and 2 level questions correctly than in 2012. Similarly in grade 4 there was a drop in addition and subtraction, multiplication and division, and grades 3 and 4 questions answered correctly.

Figure 7. Kwara: Average scores in test sub-scales, CS1 and CS2



4 Conclusion and implications of survey findings for Kwara's ESSPIN programme

Between 2012 and 2014, the average levels of output, outcome and impact indicators for Kwara's schools improved for SBMC functionality and inclusion of women and children, school quality and literacy pupil test scores. However, there was no change in teacher competence or headteacher effectiveness, and schools appeared to have become less inclusive. Schools which had benefited from ESSPIN's Output stream 4 support tended to perform better in inclusiveness, SBMC indicators and overall school quality. Teachers who received training during the initial phase of ESSPIN's Output 3 activities, during 2009–2011, were not significantly better than other teachers in Kwara in 2014.

The school improvement (Output stream 3) component of ESSPIN is aimed at creating better learning opportunities for children by improving the quality of schools. In Kwara, all government primary schools received the same support under the Output stream 3, leaving no comparison group to see whether schools improved more or less when varying the level of intervention. Some of the indicators targeted by Output stream 3 did not see the improvements which might be expected: there was little change in teacher effectiveness or headteacher competence, and then some positive and some negative changes to pupil test scores. This stagnation may reflect the fact that the teacher and leadership training was conducted in 2009/10 and 2010/11, and that state support was not sustained after 2011/12, and so any significant change in improvement may well have occurred before the first Composite Survey (2012) and not between CS1 and CS2.

Kwara did see more success in improving SBMC functionality and the inclusiveness of women and children in SBMCs. These three indicators all improved between CS1 and CS2. This could be due to the more recent focus on Output stream 4 interventions with the aim of ensuring that the gains from school improvement are made in an inclusive manner. Schools which had received the maximum support under Output stream 4 performed better on these indicators than schools with little or no support. In fact often the maximum-intervention schools were meeting the majority of indicators in 2012, which gave them less room for improvement between 2012 and 2014 on the indicators measured here. Meanwhile the no/minimum-intervention schools started from a lower base and had further to catch up. In this case, it is perhaps unsurprising to find that schools receiving the maximum intervention did not improve any faster on these indicators than other schools which were catching up. The overall level of improvement – which reflects some improvement in the maximum intervention schools and a larger improvement in the schools with less intervention – is substantial in Kwara.

The next Composite Survey, due in 2016, will allow us to see whether the improvements in SBMC functionality and inclusiveness are sustained and if they, and other indicators of school quality, have helped improve pupil test scores in Kwara's schools.

References

- ESSPIN (2009) School Based Management Committees in Policy and Practice, Volume 1: Research Synthesis Report. Report number ESSPIN 404
- ESSPIN (2013a) Overall findings and technical report of ESSPIN composite survey 1 (2012). Report number ESSPIN 060
- ESSPIN (2013b) Kwara state report of the ESSPIN composite survey 1 (2012). Report number KW205.
- ESSPIN (2013c) Extension of the Education Sector Support Programme in Nigeria, August 2014 – January 2017 [business case for the UK Department for International Development]
- ESSPIN (2014) SSO Synthesis Report 2014
- Megill, D.J. (2014) Final Sample Design and Weighting Procedures for the ESSPIN Second Composite Survey (CS2) in Nigeria. Unpublished note written for ESSPIN and Oxford Policy Management
- Raleigh, C., Linke, A., Hegre, H., and Karlsen, J. (2010) Introducing ACLED-Armed Conflict Location and Event Data. *Journal of Peace Research*, vol. 47, no. 5, 1–10. ACLED Version 5 (1997–2014) data accessed 28 January 2015 www.acleddata.com/
- RTI International (2014) Nigeria reading and access research and activity: adaptation of Education Sector Support Programme in Nigeria's teacher capacity development strategy. Document produced for the United States Agency on International Development

Annex A ESSPIN Output stream 3 interventions

The table below shows the ESSPIN Output stream 3 interventions delivered to date in Kwara. All schools in Kwara were categorised as receiving a medium level of intervention, according to the number of years of continuous intervention.

| Expected impact | Number of schools | 2009/10 | | | 2010/11 | | | 2011/12 | | | CS1 | 2012/13 | | | 2013/14 | | | CS2 |
|-----------------|-------------------|---------|---|----|---------|---|----|---------|---|----|-----|---------|----|----|---------|----|----|-----|
| | | L | T | SV | L | T | SV | L | T | SV | | L | T | SV | L | T | SV | |
| medium | 1,485 | 6 | 3 | 30 | 6 | 3 | 30 | | | 30 | | | 30 | 6 | 3 | 30 | | |

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

Annex B ESSPIN Output stream 4 interventions

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

| Expected impact of Output stream 4 | 2010/11 | | | 2011/12 | | | CS1 | 2012/13 | | | 2013/14 | | | CS2 | Level of output stream 3 intervention |
|------------------------------------|---------|---|---|---------|---|---|-----|---------|---|---|---------|----|---|-----|---------------------------------------|
| | S | P | M | S | P | M | | S | P | M | S | P | M | | |
| None / control | | | | | | | | | | | | | | | med |
| Medium | | | | | | | | | | 7 | | | | | med |
| medium | | | | | | | | | | 4 | | 2 | | | med |
| maximum | | | | 7 | | 4 | | r | 4 | | 6 | 4* | | | med |

Note: S = SBMC training; P = women and children participation training; M = mentoring visits; r = one-day refresher. * Mentoring visits were by civil society-government partnership teams, except those marked with an asterisk, which were by social mobilisation officers. Difference in differences analysis using regression

Annex C Difference in differences analysis using regression

This annex presents the results of regression analysis of difference in differences for the indicators above, to compare change over time between subgroups.

The analysis in Chapter 3 above presents the difference in differences of indicator means for sub-categories of groups.

Sometimes we want to use all of the available information and compare schools which have had more or less ESSPIN intervention – a continuous scale – rather than dividing them into some or none. In this case we can use regression analysis – a statistical process for estimating relationships among variables. We model the outcome indicator as depending on time (the round of the survey, CS2 versus CS1), the intensity of intervention, and a treatment effect which is the interaction between time and intensity of intervention. The treatment effect tells us if an increase in the level of intervention increased the speed at which the outcome improved. Regression results are reported as a series of 'coefficients' – numbers representing the strength of the relationship with the outcome of interest.

| Coefficient | Meaning of coefficient if positive and significant |
|-------------------|--------------------------------------------------------------------------------------------------------------------------------|
| Time (CS2 v. CS1) | the outcome improved over time |
| Intervention | the higher the level of intervention, the more effective (regardless of change over time) |
| Treatment | the higher the level of intervention, the more or faster the outcome improved over time – this is our key indicator of success |

School inclusiveness

A regression on the inclusiveness score has been used for difference in difference analysis of schools in Kwara depending on the level of ESSPIN intervention (Table 30). We regress the inclusiveness score on the round of the survey, the level of Output stream 4 intervention and a 'treatment effect', which tells us whether schools with more intervention changed more between 2012 and 2014 than schools with less intervention. The results suggest that all schools – those expecting an improvement due to ESSPIN and those not – became less inclusive between 2012 and 2014. This is shown by the negative time coefficient. However, the positive treatment effect suggests that schools which received more ESSPIN Output stream 4 support worsened less than other schools (although these results are not statistically significant). In other words, the ESSPIN support helped counteract a drop in inclusiveness.

Table 30. Kwara: School inclusiveness difference in differences (regression)

| Regression on inclusiveness score | | | |
|-----------------------------------|-------------|--------|---|
| Time (CS2 v. CS1) | coefficient | -10.32 | * |
| | SE | 1.6 | |
| Intervention | coefficient | -3.93 | |
| | SE | 2.33 | |
| Treatment | coefficient | 4.08 | |
| | SE | 2.38 | |
| | N | 204 | |

Note: * indicates a significant coefficient (p < .05).

SBMC functionality and inclusiveness

A regression to assess the impact of ESSPIN support on the change in SBMC functionality finds a positive and significant time effect, showing that all schools improved their SBMC functionality (Table 31). The intervention effect is also significant, showing that schools with more ESSPIN intervention had more functional SBMCs. However, the treatment effects were zero – suggesting that receiving more support does not lead to a faster improvement in SBMC functionality.

Table 31. Kwara: SBMC functionality difference in differences (regression)

| Regression on number of SBMC functionality criteria met | | | |
|---------------------------------------------------------|-------------|-------|---|
| Time (CS2 v. CS1) | coefficient | 3.01 | * |
| | SE | 0.52 | |
| Intervention | coefficient | 3.50 | * |
| | SE | 0.64 | |
| Treatment | coefficient | -0.97 | |
| | SE | 0.75 | |
| | N | 203 | |

Note: * indicates a significant coefficient ($p < .05$).

The regression on women's inclusiveness in SBMCs also finds that although all schools improved in women's inclusiveness over the two years, those with no or minimal intervention actually improved more than those with maximum intervention (Table 32). This is shown in the negative treatment effect, suggesting that the more ESSPIN support received, the slower the SBMC improved in women's inclusiveness

Table 32. Kwara: SBMC women's inclusiveness difference in differences (regression)

| Regression on number of women's inclusiveness criteria met | | | |
|------------------------------------------------------------|-------------|-------|---|
| Time (CS2 v. CS1) | Coefficient | 1.74 | * |
| | SE | 0.3 | |
| Intervention | coefficient | 1.89 | * |
| | SE | 0.25 | |
| Treatment | coefficient | -0.86 | * |
| | SE | 0.34 | |
| | N | 201 | |

Note: * indicates a significant coefficient ($p < .05$); schools that did not have SBMCs at all in CS1 are excluded from the analysis.

A regression analysis of difference in differences in SBMC children's inclusiveness finds that all schools improved over time and that schools which received more ESSPIN support were generally more inclusive of children. However, this second group were no faster at improving between 2012 and 2014 than other schools (Table 33).

Table 33. Kwara: SBMC children's inclusiveness difference in differences (regression)

| Regression on number of children's inclusiveness criteria met | | | |
|---------------------------------------------------------------|-------------|-------|---|
| Time (CS2 v. CS1) | coefficient | 0.81 | * |
| | SE | 0.21 | |
| Intervention | coefficient | 1.79 | * |
| | SE | 0.3 | |
| Treatment | coefficient | -0.33 | |
| | SE | 0.42 | |
| | N | 204 | |

Note: * indicates a significant coefficient ($p < .05$); schools that did not have SBMCs at all in CS1 are excluded from the analysis.

School quality

The regression method for difference in differences in school quality finds that generally all schools improved their school quality score between 2012 and 2014, as shown by the positive and significant time effect (Table 34). The coefficient on intervention is also found to be significant, meaning that if schools have received more ESSPIN Output stream 4 support they tended to have a higher quality score in both 2012 and 2014. However, the treatment effect is not significant, so we cannot say that the more ESSPIN support a school received, the faster it improved its quality score between 2012 and 2014.

Table 34. Kwara: School quality difference in differences (regression with continuous intervention variable)

| Regression on school quality score | | | |
|------------------------------------|-------------|-------|---|
| Time (CS2 v. CS1) | coefficient | 12.12 | * |
| | SE | 3.61 | |
| Intervention | coefficient | 16.15 | * |
| | SE | 4.09 | |
| Treatment | coefficient | -6.24 | |
| | SE | 4.88 | |
| | N | 172 | |

Note: * indicates a significant coefficient ($p < .05$).

Pupil learning

In terms of pupil learning achievement, the regression analysis finds that generally over time test scores in the literacy tests increased in all schools, as shown by the positive significant time coefficients (Table 35). The intervention coefficients are not found to be significant. The treatment coefficients are negative in all the tests, and significantly so in grade 2 numeracy. This suggests that scores actually improved more slowly (or in fact worsened faster) in schools which had benefited from more intervention over the period.

Table 35. Kwara: Pupil test score difference in differences (regression)

| Regression on pupil test score | | L2 | | L4 | | N2 | | N4 | |
|--------------------------------|-------------|-------|---|-------|---|-------|---|-------|--|
| Time (CS2 v. CS1) | coefficient | 7.86 | * | 7.11 | * | -4.76 | | -3.28 | |
| | SE | 2.53 | | 2.84 | | 2.53 | | 2.25 | |
| Intervention | coefficient | -2.92 | | -3.42 | | 1.36 | | 5.14 | |
| | SE | 3.44 | | 4.06 | | 2.76 | | 3.29 | |
| Treatment | coefficient | -3.29 | | -3.97 | | -7.5 | * | -4.39 | |
| | SE | 3.9 | | 3.74 | | 3.61 | | 3.76 | |
| | N | 203 | | 203 | | 203 | | 202 | |

Note: * indicates a significant coefficient ($p < .05$).