Oxford Policy Management

ESSPIN Composite Survey 2

Gender and inclusion report

Sourovi De

Stuart Cameron

May 2015

Acknowledgements

The authors are grateful to a number of people who have helped in producing this report. 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, Sandra Graham, Fatima Aboki, Manjola Kola, Sarah Amahson, Helen Pinnock, Jake Ross, Simon Thomson and Kayode Sanni reviewed the initial concept note and provided valuable comments on the drafts of this report. 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 head teachers, 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
Registered in England: 3122495

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

Executive summary

This report presents findings from the first and second rounds of the Composite Survey conducted in 2012 and 2014, respectively (CS1 and CS2). The survey covered a wide range of indicators at the teacher, head teacher, school-based management committee (SBMC), and pupil levels. This report attempts to establish how inclusive practices in schools and SBMCs in Education Sector Support Programme in Nigeria (ESSPIN) states are changing over time, whether schools which receive ESSPIN's interventions are working better than those which do not (Part A), and whether there are differences in education outcomes by gender and background (Part B). The main findings are as follows:

Part A: Inclusive practices in schools and SBMCs

School inclusiveness: There is evidence of a worsening over time in school inclusiveness between CS1 and CS2 across all schools in the six programme states. However, schools which received ESSPIN intervention are significantly more inclusive than the control schools, especially in terms of activities to improve access for disadvantaged children and in the use of different assessment methods.

Head teachers' actions to improve pupil attendance: Overall, the average number of actions taken by head teachers to improve pupil attendance is lower in CS2 than in CS1 across all schools in the six programme states. We did not find significant differences between ESSPIN and non-ESSPIN schools in terms of the number of actions that they said they had taken to improve pupil attendance.

Spatial and gender inclusiveness in classrooms: Spatial inclusiveness is the extent to which teachers include children sitting in all parts of the classroom during a lesson. Overall, teachers in CS2 are, on average, less spatially inclusive than those in CS1; but in CS2, teachers who have had ESSPIN training are more spatially inclusive than those who have not. Gender inclusiveness is measured as the extent to which boys and girls participate in the lesson in equal numbers. There was no evidence of any change over time in gender inclusiveness, but there is evidence that teachers in ESSPIN schools are more inclusive of girls than teachers in other schools.

SBMC functionality: SBMC functionality improved between CS1 and CS2 across all schools in the six programme states. Across a wide range of criteria of SBMC functionality, ESSPIN schools performed better than non-ESSPIN schools. The average ESSPIN school met around five of the nine criteria, while the control schools met around two, and overall 68–70% of ESSPIN schools met the SBMC functionality standard, compared to 18% of the control schools.

SBMC women's inclusiveness: There was no significant change in the extent to which SBMCs were inclusive of women between 2012 and 2014. This is in accordance with earlier qualitative research on SBMCs, which showed that women's participation in some states and in some communities within states was 'highly constrained'. However, around half of ESSPIN schools met the SBMC women's inclusiveness criteria, compared to only 3% of non-ESSPIN schools.

SBMC children's inclusiveness: Similarly, participation of children in school management is seen as 'not accepted' due to sociocultural norms. However, there was a small but significant increase between CS1 and CS2 in the overall number of criteria met, but not in the proportion of schools meeting the standard, which remained low at 6% across all schools in the six programme states. However, there are large differences between ESSPIN and non-ESSPIN schools, with around one in five ESSPIN schools meeting our criteria for children's inclusiveness, compared to only 2% of non-ESSPIN schools. In this context, where children are not expected to attend meetings or speak

in front of their elders, it would require significant preparation and support to enable the meaningful participation of children.

Part B: Differences in education outcomes by gender and background

Gender differences: We find no significant gender differences in mean pupil test scores across all CS2 schools as well as within ESSPIN and non-ESSPIN schools.

Household wealth: Pupils from the richest 20% of households answered twice as many questions correctly as those from the poorest, and the correlation between test scores and household assets is stronger for literacy than numeracy. Does being in an ESSPIN school in CS2 dampen the effect of inequality in household assets? We find evidence that the poorest children are benefiting disproportionately from ESSPIN, and that ESSPIN schools have smaller wealth disparities than the control schools.

Remoteness of schools: Pupils in urban schools performed significantly better than those in rural schools, by factor of 1.5–2 (depending on the test). There is a rural–urban gap between both ESSPIN and non-ESSPIN schools. Pupils have lower test scores in schools that are more remote. For every increase of 10 kilometres in distance between the school and the headquarters of the local government authority, average test scores dropped by around one percentage point.

Speaking a minority language: We do not find any evidence that pupils who predominantly speak a minority language (i.e. *not* Igbo in Enugu; Hausa in Jigawa, Kano, Kaduna; and Yoruba in Kwara) attain significantly different results in the numeracy and literacy tests than those who speak the majority language of each state.

Being overage for one's grade: Around 21% of the pupils in the CS2 sample are overage for their grade. Across tests and grades, overage pupils performed better than their counterparts who were age-appropriate for their grades.

Disability: The CS2 pupil tests included a number of questions assessing various forms of physical disability among sampled pupils; children who do not have a particular ability were not made to sit through questions that required that ability. In total, 54 pupils from the CS2 sample (i.e. less than 0.5%) were found to be disabled. As these children cannot be taken as a representative sample—and were disabled in different ways—we have not analysed their test results separately.

Teacher competence: Overall, 50% of CS2 teachers were female, with wide variation between northern (Jigawa: 14%) and southern states (Lagos, Enugu 84%). Female teachers performed significantly better than their male counterparts on almost all the logframe teacher competence criteria for all six states taken together. At the state level, these gender differences are mirrored only in Enugu, Kano and Kwara. The number of female teachers who passed the teacher tests in English and mathematics is twice that of male teachers, and the number of female teachers who met the teacher competence logframe standard is three times the number of male teachers who met the same. Teachers in ESSPIN schools are more than twice as likely to meet the overall teacher competence standard than male teachers in their schools. Female teachers perform significantly better than their male counterparts on spatial inclusiveness, but we find no significant difference in gender inclusive practices between male and female teachers.

Head teacher effectiveness: Overall, 34% of the schools in our CS2 sample were headed by female head teachers, with wide variation between northern (Jigawa 1%) and southern (Lagos 86%) states. In CS2, across all six states, female head teachers appear to be performing significantly better than their male counterparts: the proportion of female head teachers meeting the effectiveness standard is more than twice that of male teachers. Similarly, in ESSPIN schools, female head teachers are more than twice as likely to meet the logframe standard for head teacher effectiveness (49%) as male head teachers (22%). We do not find these gender-based differences at the state-level (except in Enugu).

Some good, interesting and challenging news from the gender and inclusion report

Positive results in this report include:

- ESSPIN schools are more inclusive than other schools both in terms of how the school is managed (p. 21) and in terms of the participation of women and children in SBMCs (p. 28, 29).
- Teachers in ESSPIN schools are more 'spatially inclusive' than teachers in non-ESSPIN schools: children from different parts of the classroom participate more in their lessons. There is also evidence that they are also more 'gender inclusive': boys and girls participate more equally (p. 24).
- Across all schools in the six states, we find that SBMC functionality appears to have improved between CS1 and CS2, and ESSPIN schools have better functioning SBMCs (according to ESSPIN's logframe criteria) than the control schools (p. 26).
- There is evidence that the poorest children are benefiting disproportionately from ESSPIN's Output 3 interventions. Treatment schools are associated with smaller wealth disparities than the control ones (p. 36).

Some interesting findings from the report include:

- There is no evidence of gender differences in pupil test scores (p. 36).
- Pupils from the richest 20% of households answer twice as many questions correctly as those from the poorest 20% of households (p. 36).
- Pupils in urban schools performed better than those in rural schools. The rural–urban gap, though less pronounced in ESSPIN schools, is still significant (p. 39).
- Test scores are lower in more remote schools: for every increase of 10 kilometres in distance between the school and the headquarters of the local government authority, average test scores drop by around one percentage point (p. 40).
- There is no evidence that pupils who speak the majority language of their state at home score any better or worse than those who speak the minority languages (p. 40).
- Overage pupils perform better than those who were age-appropriate for their grades (p. 42).

Some challenges identified in the report include:

- Across all schools in the six states, we find evidence of significant worsening in school inclusiveness over time between CS1 and CS2 (p. 21).
- Average number of actions taken by head teachers to improve pupil attendance is lower in CS2 compared to CS1. This is across all schools in the six states (p. 23).
- Teachers sampled in CS2 schools were performing worse, on average, on spatial inclusion than in CS1 (p. 24).
- Relatively few SBMCs meet standards for women's and children's inclusiveness, and there was no significant improvement in women's inclusiveness between CS1 and CS2 (p. 28).

Table of contents

Acknowledgements i
Executive summary ii
Part A: Inclusive practices in schools and SBMCs ii
Part B: Differences in education outcomes by gender and background iii
Table of contents v
List of tables, boxes and figures vii
List of abbreviations ix
1 Introduction 10
1.1 ESSPIN's community engagement and learner participation activities 10
1.2 Coverage and scale-up since 2010–11 13
1.3 Structure of this report 14
2 Methods 15
2.1 Sample and weights 15
2.2 Training, pilots and fieldwork model 17
PART A: INCLUSIVE PRACTICES IN SCHOOLS AND SBMCs 19
3 Introduction to Part A 20
4 Inclusive practices in schools 21
4.1 School inclusiveness 21
4.2 Head teachers' actions to improve pupil attendance 23
4.3 Spatial and gender inclusiveness 24
5 SBMC functionality and inclusiveness 26
5.1 SBMC functionality 26
5.2 SBMC women's inclusiveness 28
5.3 SBMC children's inclusiveness 29
5.4 Summary of SBMC data collected by SMOs 30
PART B: DIFFERENCES IN EDUCATION OUTCOMES BY GENDER AND BACKGROUND 33
6 Introduction to Part B 34
7 Pupil learning outcomes 35
7.1 Enrolment of boys and girls 35
7.2 Gender differences 36
7.3 Wealth 36
7.4 School location 39
7.5 Speaking a minority language 40
7.6 Being overage for one's grade 42
7.7 Disability 45
8 Teacher competence: gender differences 49
8.1 Gender composition of CS2 teacher sample 49
8.2 Teacher competence criteria 49
8.3 Spatial and gender inclusion practices 51
9 Head teacher effectiveness: gender differences 52
9.1 Gender composition of head teachers in CS2 52

9.2 Head teacher effectiveness criteria 52
9.3 Actions to improve attendance 55
10 Conclusion 56
References 57
Annex A Inclusion practices in schools and SBMCs: Results by ESSPIN's Output 3 interventions 58
A.1 School inclusiveness 58
A.2 Head teachers' actions to improve pupil attendance 59
A.3 Spatial and gender inclusiveness 59
A.4 SBMC functionality 60
A.5 SBMC women's inclusiveness 60
A.6 SBMC children's inclusiveness 61
Annex B ESSPIN Output 3 intervention categories 62
Annex C ESSPIN Output 4 intervention categories 63
Annex D Overlaps between ESSPIN outputs 3 and 4 64

List of tables, boxes and figures

Table 1: Sample coverage in CS2 ... 17
Table 2: Instruments used in CS2 ... 18
Table 3: School inclusiveness in CS1 and CS2 ... 22
Table 4: School inclusiveness in CS2 by Output 4 intervention group ... 23
Table 5: Head teachers' actions to improve pupil attendance in CS1 and CS2 schools ... 24
Table 6: Head teachers' actions to improve pupil attendance in CS2 schools by Output 4 intervention groups ... 24
Table 7: Spatial and gender inclusiveness by teachers in CS1 and CS2 schools ... 25
Table 8: Spatial and gender inclusiveness (CS2) by Output 4 intervention groups ... 25
Table 9: SBMC functionality in CS1 and CS2 ... 27
Table 10: SBMC functionality in CS2 by Output 4 intervention groups ... 27
Table 11: SBMC's women's inclusiveness in CS1 and CS2 ... 28
Table 12: SBMC women's inclusiveness in CS2 by Output 4 intervention group ... 29
Table 13: SBMC children's inclusiveness in CS1 and CS2 ... 29
Table 14: SBMC children's inclusiveness in CS2, between Output 4 intervention groups ... 30
Table 15: SMO reporting: SBMC performance on key indicators ... 31
Table 16: Enrolment of girls and boys in primary grades of government schools, 2009–13 ... 35
Table 17: Gender disaggregated pupil learning outcomes in CS2 ... 36
Table 18: Gender-disaggregated pupil learning outcomes in CS2 by intervention groups ... 36
Table 19: CS2 pupil test scores in rural and urban schools, by ESSPIN status ... 40
Table 20: Change in test scores per 10-kilometre increase in distance from school to LGA headquarters 40
Table 21: Main language spoken by pupils at home in CS2, by state ... 41
Table 22: Pupil test scores by language status in CS2 ... 41
Table 23: Pupil test scores in CS2, by language status and Output 3 intervention group ... 42
Table 24: Proportion of overage pupils in CS2 overall and by state ... 43
Table 25: Extent of overage among CS2 pupils (% of overage pupils) ... 43
Table 26: Mean scores across tests in all CS2 schools by overage status ... 43
Table 27: Mean scores across tests in all CS2 schools by overage status and intervention group 44
Table 29: Descriptive statistics on disabled pupils from CS2 ... 47
Table 30: Percentage of public primary school teachers who are female ... 49
Table 31: Teacher competence in CS2 by gender groups ... 50
Table 32: Teacher competence in CS2, by gender and Output 3 intervention groups ... 51
Table 33: Spatial and gender inclusiveness by teachers in CS2 schools, by gender groups ... 51
Table 34: Percentage of schools headed by female head teachers in CS2 sample ... 52
Table 35: Head teacher effectiveness in CS2 schools, by gender ... 54
Table 36: Head teacher effectiveness in CS2 schools, by intervention categories and gender 54
Table 37: Head teachers' actions to improve pupil attendance in CS2, by gender ... 55
Table 38: School inclusiveness in CS2 by Output 3 intervention group ... 58
Table 39: Head teachers' actions to improve pupil attendance in CS2 schools by Output 3 intervention groups ... 59
Table 40: Spatial and gender inclusiveness by teachers in CS2 schools by Output 3 intervention group ... 59
Table 41: SBMC functionality in CS2 by Output 3 intervention groups ... 60
Table 42: SBMC women's inclusiveness in CS2 by Output 3 intervention group ... 60
Table 43: SBMC children's inclusiveness in CS2 between Output 3 intervention groups ... 61
Table 44: Overlap between outputs 3 and 4 ... 64

Box 1: Standard for school inclusiveness (meeting needs of all pupils) ... 21
Box 2: Logframe standard for SBMC functionality ... 26
Box 3: Logframe standard for SBMC women's inclusiveness ... 28
Box 4: Logframe standard for SBMC children's inclusiveness ... 29
Box 5: Question in CS2 on pupil household wealth ... 37

Box 6: Question in CS2 on main language spoken at home by pupils 40
Box 7: Instructions to data collectors for screening hearing-impaired pupils 46
Box 8: Instructions to data collectors for screening pupils unable to speak (numeracy test).... 46
Box 9: Instructions to data collectors for screening pupils unable to speak (literacy test)..... 46
Box 10: Instructions to data collectors for screening visually impaired pupils..... 46
Box 11: Instructions to data collectors for screening pupils unable to speak (numeracy test).... 47
Box 12: Logframe standard for teacher competence 50
Box 13: Logframe standard for head teacher effectiveness 53

Figure 1: ESSPIN's community engagement and learner participation interventions..... 12
Figure 2: Number of schools receiving ESSPIN Output 4 interventions 13
Figure 3: Number of schools receiving various interventions under ESSPIN Output 4 14
Figure 4: Average scores in grade 4 numeracy and literacy tests for all CS2 schools by household asset quintile..... 37
Figure 5: Does being in an ESSPIN school reduce the effect of wealth inequality?..... 38
Figure 6: CS2 Pupil test scores in rural and urban schools 39

List of abbreviations

ASC	Annual school census
CAPI	Computer-assisted personal interviewing
CS1	Composite survey 1 (2012)
CS2	Composite survey 2 (2014)
CSO	Civil society organisations
EPDC	Education Policy and Data Centre
ESSPIN	Education Sector Support Programme in Nigeria
L2	Grade 2 literacy test
L4	Grade 4 literacy test
LGA	Local government area
LGEA	Local government education authority
MLA	Measurement of Learning Achievement
N2	Grade 2 numeracy test
N4	Grade 4 numeracy test
OPM	Oxford Policy Management
pp.	percentage point
SBMC	School-based management committee
SDP	School Development Plan
SIP	School Improvement Programme
SMO	Social mobilisation officer
SUBEB	State Universal Basic Education Board
WG	UN Statistical Commission's Washington Group on Disability Statistics

1 Introduction

The aims of CS1 and CS2 are to assess the effects of the states' integrated School Improvement Programme (SIP) and to report on the quality of education in the six states supported by ESSPIN. The surveys address five output indicators: teacher competence, head teacher effectiveness, school development planning, school-based management committee functionality, and inclusive practices in schools. It also addresses one outcome indicator (school quality) and one impact indicator (pupil learning achievement). In particular, CS2, conducted in 2014, aims to provide post-intervention data which can be compared to data from the first round of the survey collected in 2012, in order to evaluate the extent of improvement in key indicators and gauge programme success.

The overall CS2 report (Cameron, 2015) presents findings from the CS2 and comparisons between CS1 and CS2, covering all of ESSPIN's output, outcome and impact indicators, while a related set of reports discusses results for each of the six states.

This gender and inclusion report, therefore, is largely descriptive and focuses on a sub-set of indicators. Based on cross-sectional analysis undertaken for this report, all results discussed below are associational (and not causal) and therefore do not conclusively refer to programme effectiveness.[1] In particular, attempts have been made in this report to understand:

- in Part A, whether inclusive practices in schools and SBMCs in programme states are changing over time and whether schools which receive ESSPIN interventions aimed at community engagement and learner participation (i.e. ESSPIN Output 4 interventions) are working better than those which do not; and,
- in Part B, whether there are differences in education outcomes by gender and background overall, and between schools which received the SIP (i.e. ESSPIN Output 3 interventions).

1.1 ESSPIN's community engagement and learner participation activities

ESSPIN aims to bring about better learning outcomes for children of basic education school-age in six states, with a range of activities at the state, national, local and school levels. It has four output streams, focusing on (i) strengthening federal government systems, (ii) increasing the capability of state and local governments in governance and the 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 (DFID, 2013).

As Figure 1 shows, under the fourth of these outputs (circled in red below), ESSPIN's community engagement and learner participation activities (Output 4) aim to increase inclusive practices in basic education to enhance access to and quality of basic education, and therefore to improve learning outcomes for children in its six states—Enugu, Jigawa, Kaduna, Kano, Kwara and Lagos. Output 4 seeks to strengthen community engagement in school improvement and wider access by ensuring:

- SBMCs function according to their roles and responsibilities set out in state policy guidelines, and women and children influence the way schools are run and play a role in school improvement;
- Community and government organisations are better able to press for school improvement; and

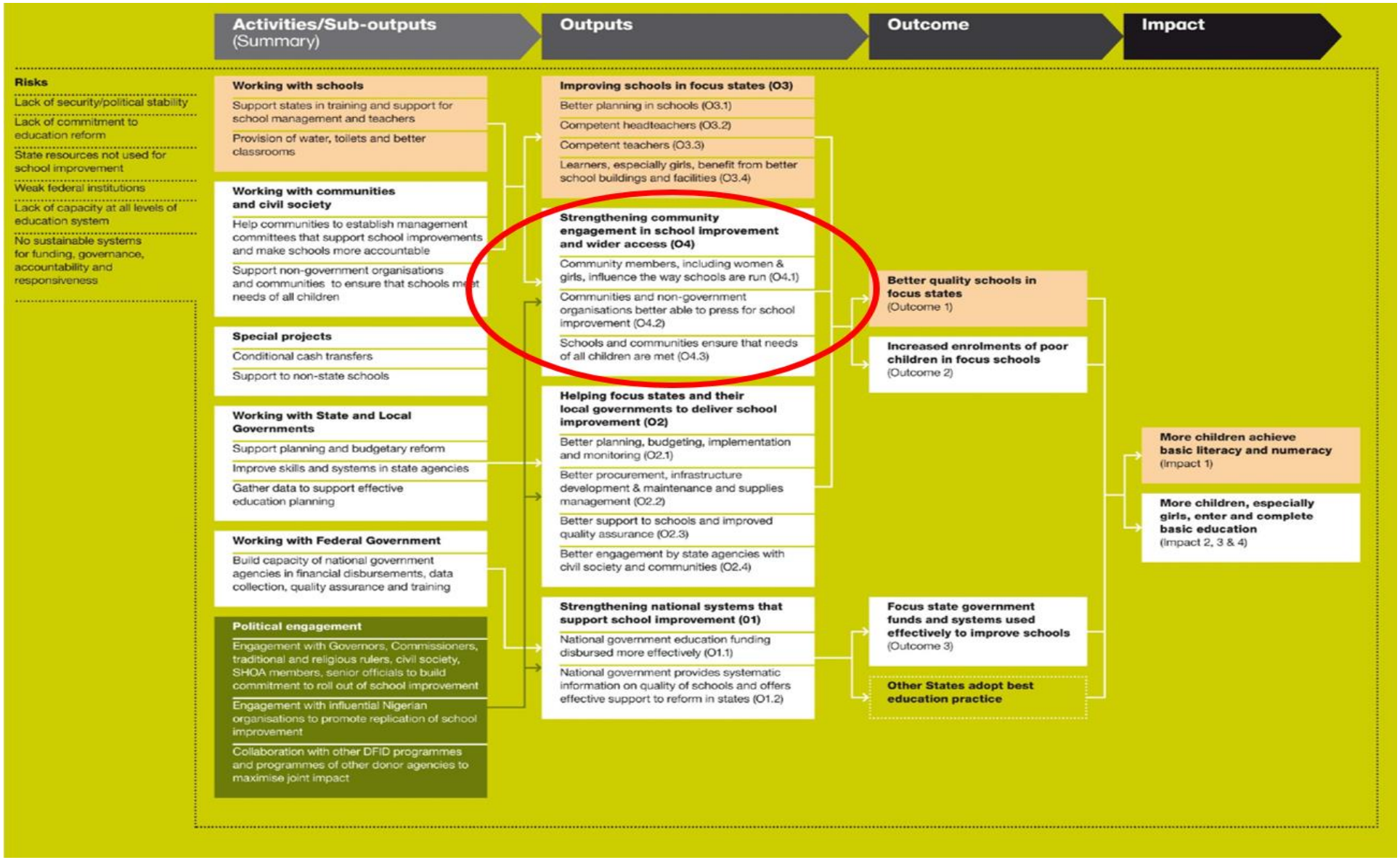
[1] The overall report (Cameron, 2015) can be referred to for further analyses based on difference-in-difference and matching methods.

- Schools and communities support inclusive education.

At the school and community levels, activities under Output 4 typically include activation and development of functioning SBMCs, which then act as the key link between the school and the community, including women, children (both enrolled and out-of-school), girls, and nomadic communities. SBMC activation and training (run for seven days for 15–17 members—men, women and children) is accompanied by more specific training to enhance the participation of women and children (six days). This is followed by eight days of mentoring visits by civil society organisations (CSOs) and social mobilisation officers (SMOs) over a period of 18 months in each school-community to provide additional training, support SBMCs on the job, and monitor progress against key criteria.

Among key results of Output 4, close to 6,000 schools received one or more activities under Output 4 in 2013–14 (whether SBMC training/refresher, women's and children's participation training, or mentoring visits), compared to 360 schools when Output 4 first started in 2010–11. It is expected that 4,500 functioning SBMCs will be achieved by 2017, and that the number of schools demonstrating inclusive practices will have reached 7,000 by then, compared to 1,600 in 2014 (DFID, 2013). Output 4 activities are assessed annually by self-assessment by SMOs and are therefore not included in the Composite Survey. It also involves support for Islamiyya, Quranic, and Tsangaya Education schools and for nomadic schools in Jigawa State, but detailed analysis of these Output 4 activities is beyond the scope of the school-based Composite Surveys.

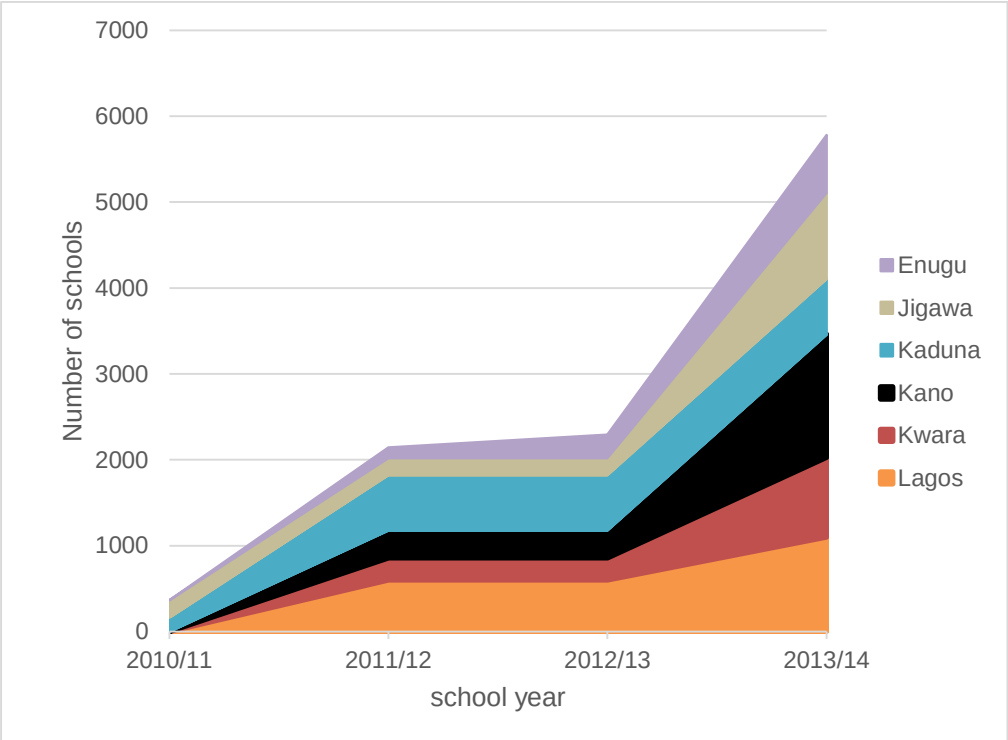
Figure 1: ESSPIN's community engagement and learner participation interventions



1.2 Coverage and scale-up since 2010–11

ESSPIN's activities under Output 4 have seen significant scaling-up, particularly during the 2013–14 school year (Figure 2). While by the end of 2012–13, 15% of all schools[2] in the six states had been in receipt of one or more interventions under Output 4, this rose to 37% in 2013–14. However, these overall coverage rates conceal large differences across states. In Lagos the intervention was rolled out to all schools in 2013–14;[3] in Kaduna, 15% schools were covered initially, and this has remained largely constant since 2012–13.

Figure 2: Number of schools receiving ESSPIN Output 4 interventions



Source: authors' calculations based on intervention information provided by ESSPIN.

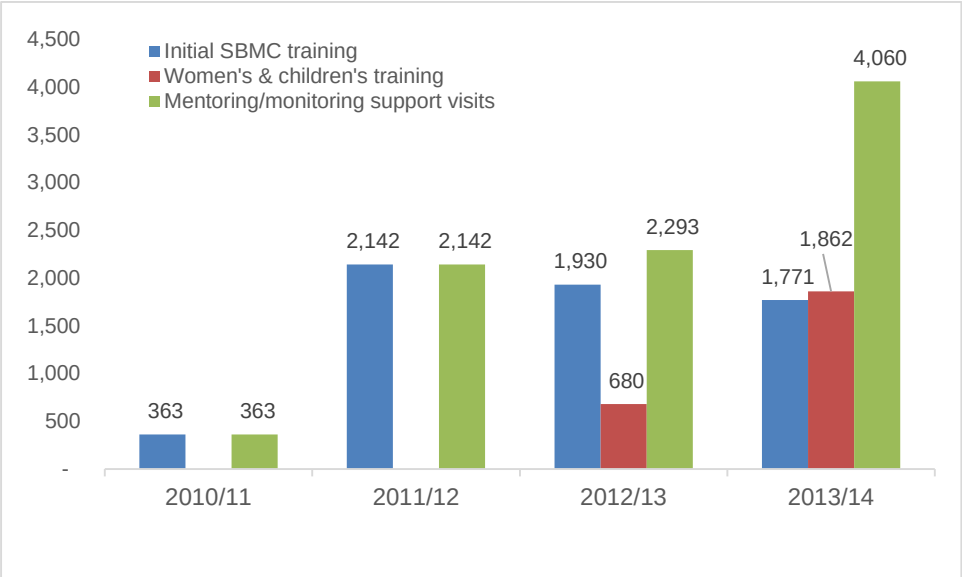
Figure 2 above also masks differences in coverage and scaling-up across various Output 4 activities. As Figure 3 below shows, while activities related to SBMC and mentoring visits started in 2010–11 and were scaled up massively in 2011–12, women's and children's participation and training activities started only in 2012–13 and have been scaled up rapidly since. This bears testimony to the model of SBMC development, in which the kind of training received by schools changes over time. In other words, the model first activates SBMCs and trains them initially, and then provides them with ongoing mentoring support (a minimum of eight school-level visits across an 18-month period). This also involves monitoring their progress by SMOs and offering refresher/consolidation support in subsequent years.

[2] We use the schools in the 2012–13 Annual School Census (ASC) as the denominator for calculating the proportions of schools that received intervention. There have been some changes from year to year to the total number of schools included in the census, but these are unlikely to have been large enough to make a large difference to our estimates of ESSPIN coverage.

[3] Lagos is a unique case in that it completely restructured its SBMC development model over the last year from a cluster-based system to the ESSPIN-supported, school-based system.

In short, the model starts with intense training, which reduces and levels off over time: the initial seven-day training will not usually be repeated (but every year SBMCs continue to receive mentoring visits). Hence, in Figure 3 we see a significant growth in mentoring visits and a gradual reduction in the number of days of initial SBMC training. When all schools have been reached, we would expect the number of mentoring visits remain roughly constant in future years, and the number of training days to reach a relatively low but stable number.

Figure 3: Number of schools receiving various interventions under ESSPIN Output 4



Source: authors' calculations based on intervention information provided by ESSPIN.

In sum, results presented in this report should be interpreted in the light of the discussion above on coverage and scaling up of Output 4 intervention. A school which only recently started receiving Output 4 interventions in the school year 2013–14 as part of the scale-up will be unlikely to show the expected outcome and impact. This is because CS2 was conducted in May–June 2014—towards the end of the 2013–14 school year—which we judge to be too soon for the additional intervention to have taken effect in ways that can be measured with most of our indicators.

1.3 Structure of this report

The rest of this report is organised as follows. Section 2 describes the sample and its coverage, CS2 survey implementation, instruments and intervention categories. Then in Part A we focus on inclusive practices in schools, classrooms and SBMCs (Section 3); in particular, we discuss levels of school inclusiveness, the actions of head teachers to improve pupil attendance, and the spatial and gender inclusion behaviours of teachers. In Part A, we also discuss the extent to which SBMCs are functional and inclusive of women and children (Section 5). We summarise key findings from a recent report on data collected by SMOs to monitor SBMC development and performance (separately from the CS2 survey).

In Part B, we discuss differences in education outcomes pertaining to pupils, teachers and head teachers by gender and background (Sections 6). On pupil learning levels (Section 7), we discuss differences by gender, household wealth, remoteness of schools, language, age and disability. On teacher competence (Section 8), we discuss differences in terms of gender and, in particular, teachers' spatial and gender inclusion behaviour in the classroom. In Section 0, gender differences in head teacher effectiveness are discussed. Section 10 provides a conclusion. Annex A discusses more detailed results on inclusive practices in schools, classrooms and SBMCs. Annex B and Annex C provide details on outputs 3 and 4 intervention categories used for the analysis underlying this report.

2 Methods

2.1 Sample and weights

2.1.1 Sample design

The aim of the sample design for CS2 was to allow follow-up on schools already sampled in CS1, and to allow inferences to be drawn about what is happening in the population of schools across the six states and within each state, through the use of sample weights. The sample design prioritised the ability to draw conclusions across the six states, conceding that it would not always be possible to obtain statistically significant estimates within each state, given the high degree of variability in the types of schools that are found in some of the states, which makes it difficult to construct a representative sample.

A stratified sampling design was used for CS1. The survey covered primary schools in the six ESSPIN states—Enugu, Jigawa, Kaduna, Kano and Lagos. The sampling frame was compiled using the annual school census (ASC) with stratification by ESSPIN phases and whether or not the schools had participated in an earlier attempt to measure learning outcomes, the 2010 Measurement of Learning Achievement (MLA) exercise. The sample design took account of MLA participation in order to allow comparisons between that 2010 round of data collection and the CS1.

Stratifying by MLA participation, however, resulted in high variability in sampling weights and contributed to problems in interpreting some of the results in CS1. There were, in any case, concerns about data quality in the 2010 MLA, and so the attempt to compare with the 2010 data was dropped for CS2. In line with the recommendations of the CS2 sample design report (Megill, 2014), the school sample in two of the states (Kano and Kaduna) was increased in order to reduce problems of high variability in weighted results; and in a third state (Enugu), for which fewer schools were sampled in CS1, the sample was increased to the same level as the other states.

Teachers were randomly sampled within selected schools in both CS1 and CS2. Following another recommendation from the sampling report, it was decided to reduce the number of teachers sampled per school, from 10 in CS1 to six in CS2, partly because sampling 10 teachers per school resulted in difficulties in detecting significant effects in some states,[4] and also because there are many schools which have fewer than 10 teachers. Although it would have been useful to have tracked individual teachers, this was not seen as feasible given the limited collection of identifying information during CS1. Teachers within each school were sampled from the population present in the school on the day of the survey visit and from those teaching grades 1–6 in the present term, using the school's teacher attendance register. Fieldwork teams asked head-teachers to complete such a register in cases where it had not already been done for the day of the visit. Team supervisors entered the number of eligible teachers into the computer-assisted personal interviewing (CAPI) system, which would then randomly select 6 to be sampled.

Pupils were sampled from the pupil registers for grade 2 and 4 classes. The sample size of four pupils per test per school remained the same between CS1 and CS2. In cases where the registers were not available or had been found to be inaccurate, data collectors went to the classrooms and counted pupils instead. Again, the CAPI system randomly assigned pupils to be sampled for each

[4] In complex survey designs, clustering of results can make it difficult to obtain precise estimates. A high degree of similarity among teachers within schools, combined with a high degree of variation between schools, resulted in clustering effects in the teacher indicators. This, along with the type of sampling weights that were applied, resulted in high 'design effects' and a loss of precision in the CS1 analysis for some states. See ESSPIN (2013) and Megill (2014) for a full explanation.

test (grade 2 literacy, grade 2 numeracy, grade 4 literacy and grade 4 numeracy) from the numbers present in grades 2 and 4. As with teachers, it would have been useful to trace the same pupils over time, but this was not seen as feasible because for the children sampled in CS1, we have only their names, which is not always sufficient to identify the same children two years later. We therefore collected a random sample within each school in both CS1 and CS2.

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

A number of schools were found to operate double shifts, with some classes taught in the morning and others in the afternoon. Where necessary, the previous afternoon's register was added to the morning register for the day of the visit, in order to ensure that teachers and pupils were included whether they attended the morning or afternoon session.

2.1.2 Weights

Simple averages of the results from the Composite Survey data would not be representative of what is happening across the state, because the profile of schools in the survey is not identical to the profile of schools in the state as a whole. We overcame this problem 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. A smoothing technique was also applied to reduce the variability of the weights and avoid the statistical problems encountered in the CS1 analysis (see above and Megill, 2014). Most of the following analysis applies weights to sample statistics calculated within each round and intervention group, which can then be used as estimates of the whole population of schools in the six ESSPIN states.

2.1.3 Sample coverage

The effective sample from each state is listed in Table 1. The intended sample for CS2 consisted of 735 schools. During fieldwork, six schools in the sample were found to be ineligible for the survey, either because they were purely private Quranic schools with no non-religious education, or (in one case) because it was a special school for deaf students using different methods from regular government schools. One school (in Jigawa) was not visited because of security concerns.

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). Overall, complete test data was gathered in CS2 for 95% of the maximum possible number of students, a similar proportion to CS1. Interviews were gathered and lesson observations conducted for 80% and 78%, respectively, of the maximum possible number of teachers. This figure was relatively low due to there being fewer than six teachers in a large number of schools.

Table 1: Sample coverage in CS2

	Schools		Teachers			Pupil tests			
	Intended sample	Actual	Interview	Lessons observed	Tests	L2	L4	N2	N4
Enugu	105	105	532	519	494	388	395	384	382
Jigawa	105	103	430	425	415	393	399	396	398
Kaduna	140	140	638	625	586	539	521	538	517
Kano	175	170	773	764	645	667	635	669	632
Kwara	105	105	545	538	483	394	393	397	391
Lagos	105	105	569	545	526	415	410	413	409
Total	735	728	3487	3416	3150	2796	2753	2797	2729

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.

2.2 Training, pilots and fieldwork model

Fieldwork for CS2, including the pupil tests, was conducted using CAPI during May–July 2014. Children were given a printed pupil book to read and write on. The interviewers held a tablet computer, prompting them with the questions to be asked orally to the children, giving instructions on the administration of the different test items, including timing, and allowing them to input whether each part of each question was answered correctly or incorrectly (or not attempted at all) by the pupil.

The instruments were piloted during March 2014 after training state coordinators and monitoring officers. Pilots were conducted in Nassarawa, Kaduna and Abuja with the coordinators and monitoring officers collecting the data, and using both paper and CAPI. In total there were eight pilot days with additional piloting of revised pupil tests in Abuja. Instruments were revised through consultation with ESSPIN and state coordinators.

Table 2 lists the instruments used in CS2, together with the indicators relevant to outcomes, outputs or impact that were gathered from each instrument. The instruments were also used to gather intervention information, such as whether individual teachers had received ESSPIN training or not, and pupil-level information on socioeconomic status, age, language spoken at home, and gender. The data gathered in general allows more detailed analysis than that presented in this report, some of which is presented in the six state-level reports and the gender and inclusion report that will accompany this one. The data will also be published in anonymised form for use by ESSPIN and other researchers.

The process of revising instruments for CS2 does leave some possibility of measurement error in comparisons between CS1 and CS2. Given that training and fieldwork were extremely challenging for CS1, the priority for CS2 was to ensure consistent and manageable data collection within CS2 by setting clearer guidance for data collectors through detailed data collection manuals, greater oversight, and through a single intensive training session for all data collectors across the six states. Although we avoided large changes in instruments that would compromise comparability with CS1, any change in questionnaire format or wording, training, and data collection procedures can potentially affect the results, and this should be kept in mind. However, the change in instruments should not have affected our difference-in-differences results, provided that changes in measurement are consistent across ESSPIN and non-ESSPIN schools.

Table 2: Instruments used in CS2

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

PART A: INCLUSIVE PRACTICES IN SCHOOLS AND SBMCs

3 Introduction to Part A

This part of the report focuses on inclusive practices in schools and SBMCs in ESSPIN states. We use a wide range of indicators gathered through interviews as part of the Composite Surveys with head teachers and SBMC chairpersons, and through lesson observations, as well as referring to reports compiled by SMOs. In each case, we ask two main questions:

1. **Are schools in the six states as a whole improving over time?** We measure this by looking at the change in indicators between CS1 (2011–12 school year) and CS2 (2013–14 school year). Although it is hoped that the presence of ESSPIN in the six states will have improved school inclusiveness during this time period, this cannot be taken for granted because ESSPIN interventions have not yet been scaled up to cover most schools in the six states. Moreover, the large-scale roll-out of ESSPIN interventions during 2013–14 was unlikely to have had a major impact by the time of our survey in late 2013–14. Changes in the states as a whole reflect the presence of ESSPIN to some degree, but also factors that are beyond ESSPIN's control—such as political circumstances, conflict, and changes in pupil enrolment.

2. **Are schools receiving ESSPIN interventions on inclusiveness and community engagement (output 4) more inclusive than other schools in 2014?** We measure this by using data from CS2, and by comparing the same indicators between three groups of schools:[5]

    - *Pre-CS1 schools*: these are first phase ESSPIN pilot schools which have received the most support over time, starting from before CS1. In the pre-CS1 pilot schools, SBMCs were activated early in 2010 and have benefited from the full range of support provided by CSOs and government partners. This includes: community entry, SBMC activation, community sensitisation and community research on problems of their schools, gender and child awareness; SBMC training at cluster level; a total of 15 mentoring and monitoring support visits; and additional support to enhance the participation of women and children in school-based management and to support SBMC forums at LGEA level.

    - *Post-CS1 schools*: schools added as part of the roll-out of output 4 interventions during 2012–13 and 2013–14.

    - *Control schools*: schools which have not yet received any output 4 intervention.

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 (i.e. unlikely to have arisen by chance). This should not be taken as constituting rigorous hypothesis testing (given the large number of indicators tested) but it does provide a guide as 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.

[5] The pattern of ESSPIN's interventions under its Output 3 and Output 4 streams have differed somewhat by state. See Annex B, Annex C, and Annex D.

4 Inclusive practices in schools

In this section we look at inclusive practices in classrooms and schools. Across all schools in the six states, we find evidence of a significant drop in school inclusiveness between CS1 and CS2. But in 2014, schools with more ESSPIN intervention were more inclusive than other schools. Average number of actions taken by head teachers to improve pupil attendance is lower in CS2 compared to CS1. Teachers in CS2, on average, were performing worse on spatial inclusion than in CS1, though teachers in ESSPIN schools perform better than teachers in non-ESSPIN schools. On gender inclusion, we do not find evidence of change over time, but there is evidence that teachers in ESSPIN schools behave in more gender-inclusive ways than those in non-ESSPIN schools.

4.1 School inclusiveness

We measure school inclusiveness using a standard which depends on four criteria (Box 1).

Box 1: Standard for school inclusiveness (meeting 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) Head teacher states three or more actions[6] 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).
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).

The proportion of schools with two or more activities in the SDP aiming to improve access for disadvantaged children (criterion 2) has increased (Table 3). However, the proportion of schools meeting criteria 1 and 4 has decreased significantly. For criterion 1, the head teacher names more than three actions that he or she has taken to improve pupil attendance: the decline is very large and some degree of measurement error is suspected here, although the question format did not change between CS1 and CS2. Overall, across all schools in the six states, we do not find a significant difference between numbers of schools which meet the inclusiveness standard in CS2 vis-à-vis CS1.

[6] This was incorrectly stated as *more than three* actions in the CS1 report.

Table 3: School inclusiveness in CS1 and CS2

	CS1	CS2	
(1) three or more actions on attendance (%)	57.9	39.1	-
(2) two or more activities in SDP on access for disadvantaged children (%)	5.4	11.9	+
(3) >50% of teachers use two or more assessment methods (%)	70.7	62.3	
(4) >50% of teachers spatially inclusive and >50% are gender inclusive (%)	33.4	23.4	-
Number of inclusiveness criteria fulfilled (/4)	1.7	1.4	-
Inclusiveness score	72.2	63.7	-
School partially meets inclusiveness standard (2–4 criteria out of 4)	60.4	46.5	-
School fully meets inclusiveness standard (3–4 criteria out of 4)	18.8	12.7	

Note. + / - indicate a statistically significant positive/negative difference ($p < .05$); blank cells indicate no statistically significant difference.

Within CS2, schools with more ESSPIN intervention are more inclusive than other schools (Table 4). However, even among schools receiving inclusiveness support through Output 4 from ESSPIN, around 70% do not yet achieve enough of these criteria (3 or 4) to reach our inclusiveness standard. Schools which received ESSPIN Output 4 intervention are significantly more inclusive than the control schools in terms of activities to improve access for disadvantaged children (criterion 2), use of different assessment methods (criterion 3), and in terms of the overall inclusiveness score and the proportion of schools meeting standards. Pre-CS1 schools, i.e. schools which had started receiving interventions aimed at increasing inclusiveness and community participation before CS1, are also significantly more spatially and gender inclusive than non-ESSPIN schools. Overall 72% and 64% of ESSPIN pre-CS1 and post-CS1 schools (respectively) and only 40% of non-ESSPIN schools fully meet the school inclusiveness standard.[7] Teachers in pre-CS1 schools also used more assessment methods and involved children from more parts of the classroom in lessons.

[7] Similar results are obtained when we disaggregate the inclusion criteria by whether or not the school had received interventions under ESSPIN's output 3 (improving learning outcomes). These are presented in Annex A and discussed in the overall CS2 report (Cameron, 2015).

Table 4: School inclusiveness in CS2 by Output 4 intervention group

School inclusiveness criteria	Control	pre-CS1		post-CS1	
Inclusiveness criteria					
(1) three or more actions to improve attendance (%)	38.2	49.3		34.8	
(2) two or more activities in SDP to improve access for disadvantaged children (%)	6	23.1	+	36.9	+
(3) >50% of teachers use two or more assessment methods (%)	56.5	80.9	+	77.9	+
(4) >50% of teachers spatially inclusive and >50% are gender inclusive (%)	18.9	44.3	+	29.4	
Overall inclusiveness standard					
Number of inclusiveness criteria fulfilled (/4)	1.2	2	+	1.8	+
Inclusiveness score	61.3	69.6	+	71.1	+
School partially meets inclusiveness standard (2–4 criteria out of 4)	39.5	71.8	+	64.2	+
School fully meets inclusiveness standard (3–4 criteria out of 4) (%)	7.8	30.5	+	24.6	+
Detailed					
Number of actions to improve attendance	2.3	2.4		2.4	
Number of activities on access for disadvantaged children	0.2	0.8	+	1.3	+
Average number of assessment methods used	1.1	1.7	+	2	
Average number of zones participating in lessons	3.5	4.3	+	3.8	
Average gender equity score (0=completely unequal; 100=perfectly equal)	80.9	84.3		88.7	+

Note. + / - indicate a statistically significant positive/negative difference ($p < .05$); blank cells indicate no statistically significant difference.

4.2 Head teachers' actions to improve pupil attendance

In both rounds of the survey, head teachers in sampled schools were asked questions on actions taken, if any, to improve pupil attendance in their schools. Overall fewer categories of action[8] were taken by head teachers to improve pupil attendance in CS2 compared to CS1 (Table 5). In CS2, fewer head teachers reported the following as a means to improve pupil attendance: addressing issues such as bullying and punishment; improving the quality of teaching and learning; and implementing solutions suggested by teachers, pupils and parents. Based on the data at hand, the reasons for the deteriorating results are not entirely clear.

[8] This is different from fewer actions in a category.

Table 5: Head teachers' actions to improve pupil attendance in CS1 and CS2 schools

% of head teachers who took the following actions	CS1	CS2	
Keep up-to-date registers	45.3	44.3	
Involve SBMC in finding reasons for non-attendance	45	40.4	
Discuss reasons for non-attendance with teachers, pupils or parents	74.3	72.6	
Implement suggested solutions for non-attendance	27	7.7	-
Address issues such as bullying and punishment	22.2	3.3	-
Improve quality of teaching and learning	29.8	6.7	-
Average number of actions taken	2.75	2.31	-

Note. + / - indicate a statistically significant positive/negative difference (p < .05); blank cells indicate no statistically significant difference.

Focusing on CS2 schools (Table 6), we analyse results disaggregating by ESSPIN Output 4 phases. Actions taken by head teachers to improve pupil attendance were not significantly different in Output 4 schools. Fewer head teachers in these schools reported keeping up-to-date registers as a means to curb pupil absenteeism; more of them reported improving quality of teaching and learning. Overall, for the majority of the actions listed, the percentage of head teachers using these as a means to improve pupil attendance was not significantly different between the Output 4 and the control schools.[9]

Table 6: Head teachers' actions to improve pupil attendance in CS2 schools by Output 4 intervention groups

% of head teachers who took the following actions	Control	pre-CS1		post-CS1	
Keep up-to-date registers	50.7	33.6	-	17.7	-
Involve SBMC in finding reasons for non-attendance	40.6	34.6		44.2	
Discuss reasons for non-attendance with teachers, pupils or parents	72.7	67.2		78.3	
Implement suggested solutions for non-attendance	7	10.7		9.2	
Address issues such as bullying and punishment	3.3	5.1		1.6	
Improve quality of teaching and learning	4.3	13.6	+	15.6	+
Average number of actions taken	2.3	2.4		2.4	

Note. + / - indicate a statistically significant positive/negative difference (p < .05); blank cells indicate no statistically significant difference.

We further analyse these results disaggregated by head teacher gender in section 9 below.

4.3 Spatial and gender inclusiveness

Both CS1 and CS2 observed sampled teachers' classroom practices, including spatial and gender inclusiveness. In particular, spatial inclusiveness was included as a classroom observation instrument. Data collectors were trained to divide the classroom mentally into six zones and note the number of instances when the teacher engaged with a pupil or pupils from each of the six zones. As such, we are interested in whether the teacher engaged with at least one pupil from four different areas of the classroom during a lesson.

[9] Similar results are obtained when we break schools down according to how much intervention they had under output 3 instead of output 4 (these are reported in Annex A).

Similarly, gender inclusiveness was observed by data collectors by noting the pupil's gender each time the teacher engaged a pupil in the classroom. As such, we are interested in whether the teacher engages with boys and girls proportionally with 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 will meet the gender inclusiveness criterion.

To measure gender inclusion more precisely, we also use a continuous scale for gender equity. This is a score from 0 to 100, where 0 represents no boys participating or no girls participating, and 100 represents a situation where girls and boys participate exactly in proportion to the number of girls and boys sitting in the class.[10]

Using these definitions and indicators, we tabulate results on spatial and gender inclusiveness below. Table 7 shows that teachers sampled in CS2 schools, on average, were performing significantly worse on spatial inclusion than teachers sampled in CS1. In terms of gender inclusiveness, no significant differences were found between teachers in CS1 and CS2 schools.

Table 7: Spatial and gender inclusiveness by teachers in CS1 and CS2 schools

Indicators (%)	CS1	CS2	
Involves boys/girls proportionately (within 10%)	47.6	52.1	
Gender equity score (0=completely unequal; 100=perfectly equal)	81.6	82.5	
Involves pupils from four or more areas of the class	75.2	59.2	-
Number of zones participating in lessons (out of 6)	4.27	3.76	-

Note. + / - indicate a statistically significant positive/negative difference (p < .05); blank cells indicate no statistically significant difference.

Do teachers in ESSPIN Output 4 schools demonstrate more gender and spatial inclusiveness than their counterparts in the control schools? As Table 8 shows, teachers who are in ESSPIN schools perform significantly better than those in the control schools in terms of the spatial inclusiveness criteria.[11] They also perform better in terms of gender inclusion, although the difference is only significant when we compare the schools that started receiving output 4 interventions since 2012 (the post-CS1 schools) to the control schools.

Table 8: Spatial and gender inclusiveness (CS2) by Output 4 intervention groups

Indicators (%)	Control schools	Pre-CS1 schools		Post-CS1 schools	
Involves boys/girls proportionately (within 10%)	50.6	55.3		55.2	
Gender equity score (0=completely unequal; 100=perfectly equal)	81	84.3		87.6	+
Involves pupils from four or more areas of the class	54.8	69.6	+	67.8	+
Number of zones participating in lessons	3.6	4.2	+	4.2	+

Note. + / - indicate a statistically significant positive/negative difference (p < .05); blank cells indicate no statistically significant difference.

[10] The gender equity scale is calculated as $[100 - 100 \times abs\left(\frac{g}{g+b} - \frac{G}{G+B}\right)]$ 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.

[11] Some kind of Hawthorne effect cannot be ruled out here, in that teachers observed in the intervention schools might behave differently based on knowledge that they are part of the intervention.

5 SBMC functionality and inclusiveness

In this section, we look at SBMC functionality and inclusive practices in SBMCs. Across all schools in the six states, we find that SBMC functionality appears to have improved between CS1 and CS2. SBMCs in ESSPIN schools were also typically much more functional than those that had not received ESSPIN intervention. There was no significant change over time in women's inclusiveness between CS1 and CS2, but there may have been a modest improvement in the extent to which they were more inclusive of children. SBMCs that received support on inclusiveness from ESSPIN are much more inclusive of women and children than those that did not.

5.1 SBMC functionality

There are nine criteria in the logframe standard for SBMC functionality (Box 2).

Box 2: 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 2013–14 school year:[12]

1) Two or more SBMC meetings took place during the 2013–14 school year (written evidence);
2) SBMC conducted awareness raising activities (written or oral evidence);
3) SBMC took steps to address exclusion (written or oral evidence);
4) SBMC networked with CBOs, traditional or religious institutions, or other SBMCs (written or physical evidence);
5) SBMC interacted with LGEAs on education service delivery issues (written or physical evidence);
6) An SBMC women's committee exists (written or physical evidence);
7) An SBMC children's committee exists (written or physical evidence);
8) SBMC contributed resources for the school (written or physical evidence); and
9) SBMC chair visited the school at least three times since the start of the 2011–12 school year (written evidence).

In general, we find that SBMC functionality appears to have improved between CS1 and CS2, although the difference is not significant on every indicator (Table 9). As discussed above, these results are for all schools in the six states, covering both programme and the control schools; they thus provide an overview of the state of affairs rather than programme impact. The average school met two of the nine criteria in 2012, but in 2014 met three of the criteria—a statistically significant improvement. Most of the criteria for SBMC functionality rely on the ability to provide written or photographic evidence, or at least oral recollection of a specific event. Consequently, the criteria may reflect the quality of the record-keeping of the SBMC more than the particular aspects of functionality that they aim to measure. Examining individual criteria, SBMCs have particularly improved in terms of awareness-raising activities, networking, women's committees, and contribution of resources for the school. There was a decline in the proportion of SBMCs with written evidence of the chairperson having visited the school.

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.

[12] 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.

Both of these types of SBMC action on inclusion had been happening more often in 2013/14 than in 2011/12.

Table 9: SBMC functionality in CS1 and CS2

SBMC functionality criteria	CS1	CS2	
(1) two or more meetings this school year	29.2	27.0	
(2) conducted awareness-raising	36.0	47.9	+
(3) addressed exclusion	27.2	39.9	+
(4) networked with CBOs/institutions/other SBMCs	15.5	55.2	+
(5) interacted with LGEA	20.3	22.6	
(6) has a women's committee	13.3	26.5	+
(7) has a children's committee	19.4	21.6	
(8) contributed resources for school	40.0	54.7	+
(9) chair visited school three or more times	26.0	14.7	-
Standard G: functioning SBMC	22.1	30.7	
Number of SBMC functionality criteria met	2.3	3.3	+
Additional criteria			
Action for commonly excluded groups	13.9	24.3	+
Raised issue of children's exclusion	4.8	19.5	+

Note. + / - indicate a statistically significant positive/negative difference ($p < .05$); blank cells indicate no statistically significant difference.

Examining differences between ESSPIN Output 4 schools in CS2 vis-à-vis the control schools (Table 10), there are significant differences in the expected direction across all nine criteria. The average ESSPIN Output 4 school meet around five of the nine criteria, while the control schools meet around two—a statistically significant difference. Overall, 68% and 70% of pre- and post-CS1 schools meet the SBMC functionality standard, compared to 18% of the control schools.

Table 10: SBMC functionality in CS2 by Output 4 intervention groups

SBMC functionality criteria	Output 4 control	Pre-CS1		Post-CS1	
(1) two or more meetings this school year (%)	13.6	62.2	+	74.1	+
(2) conducted awareness-raising (%)	39.6	70.3	+	76.2	+
(3) addressed exclusion (%)	33.1	62.8	+	59.1	+
(4) networked (%)	48	76.3	+	78.2	+
(5) interacted with LGEA (%)	17.5	37.4	+	31.3	+
(6) has a women's committee (%)	15.6	61.5	+	58.0	+
(7) has a children's committee (%)	10.9	59.0	+	50.3	+
(8) contributed resources for school (%)	47.3	81.6	+	73.7	+
(9) chair visited school three or more times (%)	6.6	46.3	+	33.3	+
Standard G: functioning SBMC	18.1	68.4	+	70.4	+
Number of SBMC functionality criteria met (out of 9)	2.3	5.6	+	5.4	+
Additional criteria					
Action for commonly excluded groups	23	39	+	18	
Raised issue of children's exclusion	15.3	43.0	+	20.3	

Note. + / - indicate a statistically significant positive/negative difference ($p < .05$); blank cells indicate no statistically significant difference.

5.2 SBMC women's inclusiveness

As in the report on CS1, we also examine the extent to which SBMCs are inclusive of women's concerns. The standard on SBMC women's inclusiveness has four criteria (Box 3).

Box 3: Logframe standard for SBMC women's inclusiveness

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

1) At least one woman attended two or more SBMC meetings (written evidence);
2) A female member of SBMC raised at least one issue at SBMC meetings (written evidence or oral evidence from a female member of SBMC);
3) At least one issue raised by a female member at an SBMC meeting led to action (written, physical or oral evidence from female member of SBMC); and
4) At least one SBMC women's committee meeting took place.[13]

There was an improvement between CS1 and CS2 in the proportion of schools where a female member raised an issue during an SBMC meeting, and in the proportion with a women's committee that had met recently (Table 11). Overall, however, the change in women's inclusiveness was not large enough to be statistically significant, and the levels of women's inclusiveness continue to remain low, both in terms of average number of criteria met (out of four) and in terms of the number of schools which meet the logframe standard for women's inclusiveness in SBMCs.

Table 11: SBMC's women's inclusiveness in CS1 and CS2

SBMC women's inclusiveness criteria	CS1	CS2	Significant diff
(1) at least one woman attended two or more meetings (%)	14.8	17.1	
(2) a female member raised an issue (%)	20.3	31.9	+
(3) an issue raised by female member led to action (%)	21.5	14.8	
(4) a women's committee meets (%)	6.0	26.6	+
Number of criteria met	0.6	0.9	
Meets standard (3/4 criteria) (%)	15.4	15.8	

Note. + / - indicate a statistically significant positive/negative difference ($p < .05$); blank cells indicate no statistically significant difference.

Examining differences between ESSPIN Output 4 schools in CS2 (Table 12), there are significant differences for both types of Output 4 schools in the expected direction in all four criteria and in the overall standard. The average ESSPIN Output 4 school meets around two of the four criteria, while the control schools meet less than one on average. Overall, very few control schools are meeting the women's inclusiveness standard, but around half of ESSPIN schools are.

[13] This criterion has been slightly altered since CS1, where it was also required that the women's committee have a female leader.

Table 12: SBMC women's inclusiveness in CS2 by Output 4 intervention group

SBMC women's inclusiveness criteria	Control	Pre-CS1 schools		Post-CS1 schools	
(1) at least one woman attended two or more meetings (%)	4	50.6	+	62.7	+
(2) a female member raised an issue (%)	18.8	62.3	+	74.5	+
(3) an issue raised by female member led to action (%)	6.7	30.3	+	41.9	+
(4) a women's committee meets (%)	15.7	61.4	+	57.6	+
Number of criteria met	0.4	2.1	+	2.4	+
Meets standard (3/4 criteria) (%)	3.4	44.3	+	56.1	+

Note. + / - indicate a statistically significant positive/negative difference ($p < .05$); blank cells indicate no statistically significant difference.

5.3 SBMC children's inclusiveness

We measure SBMC children's inclusiveness using four criteria (Box 4).

Box 4: 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 last 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.[14]

There was a significant increase overall between CS1 and CS2 in the proportion of SBMCs where children raised an issue, and a small but significant increase in the overall number of criteria met, but not in the proportion of schools meeting the standard, which remained low (at 6%). As discussed above, these results are for all schools in the six states, covering both programme and control schools; it thus provides an overview of the state of affairs rather than programme impact.

Table 13: SBMC children's inclusiveness in CS1 and CS2

SBMC children's inclusiveness criteria	CS1	CS2	
(1) a child attended two or more meetings (%)	8.9	8.8	
(2) a child raised an issue (%)	10.3	20.6	+
(3) an issue raised by child led to action (%)	8.9	6.9	
(4) a children's committee meets and it has a trained facilitator (%)	1.8	14.3	+
Number of criteria met	0.3	0.5	+
Meets standard (3/4 criteria) (%)	5.7	6.2	

Note. The large increase in criterion 4 may be due to a relaxation of the evidence requirement (see footnote 14).

As with women's inclusiveness, there are large positive differences between ESSPIN and non-ESSPIN schools. Disaggregating CS2 schools by Output 4 and control schools (Table 14), we find

[14] 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 that the committee is not functioning.

large positive and significant differences on all criteria for SBMC children's inclusiveness in the expected direction. Overall, 21% and 15% of pre- and post-CS1 schools met the logframe standard for children's inclusiveness, while only 2% of the control schools did so.

Table 14: SBMC children's inclusiveness in CS2, between Output 4 intervention groups

SBMC children's inclusiveness criteria	Control	Pre-CS1		Post-CS1	
(1) a child attended two or more meetings (%)	2	30	+	26.8	+
(2) a child raised an issue (%)	11.9	43.3	+	48.1	+
(3) an issue raised by child led to action (%)	3.6	18.5	+	19.6	+
(4) a children's committee meets and it has a trained facilitator (%)	6.4	40.6	+	33.1	+
Number of criteria met	0.2	1.3	+	1.3	+
Meets standard (3/4 criteria) (%)	2	21.4	+	15.2	+

Note. + / - indicate a statistically significant positive/negative difference ($p < .05$); blank cells indicate no statistically significant difference.

5.4 Summary of SBMC data collected by SMOs

We conclude Part A with a brief discussion of the role played by SMOs in collecting data on the performance of SBMCs as a means of annually monitoring SBMC development. These are separate from the Composite Surveys, which are held every two years.

Since 2011, ESSPIN has supported the Social Mobilisation Departments of the State Universal Basic Education Boards (SUBEBs) in all six ESSPIN states to develop and institutionalise a process of reporting on SBMC development, with information flow coming up from a community and school level to state level, being summarised at local government level, and then ultimately fed into the government's planning processes.

The information collected by SMOs captures progress against key indicators of SBMC functionality (output 4 logframe indicator 1a), levels of participation of women and children in SBMCs (output 4 logframe indicator 1b) and the extent to which communities and SBMCs are supporting inclusive education (output 4 logframe indicator 3c). The reports also collect information on the enrolment of children in school and the return of children who dropped out of school as a result of SBMC action, as well as information on resources mobilised by SBMCs and communities to support children and school improvement. Here we summarise key findings from SBMC data collected by SMOs in the report released in July 2014 (ESSPIN, 2014), when data on SBMCs in 10,442 schools was collected.

A general note of caution applies to results from Enugu and Lagos:

- Enugu benefited from ESSPIN support later than all other states (starting only in 2011);
- Lagos is a unique case since it has completely restructured its SBMC development model over the last year from a cluster-based system to the ESSPIN-supported school-based system.

Summary statistics from this report are in Table 15 and brief explanations follow below. No additional/independent analysis of this data was performed by the authors.

Table 15: SMO reporting: SBMC performance on key indicators

% of schools	SBMC functionality (1a)	Women's and children's inclusiveness (1b)	Promoting inclusive education (3c)
Kwara	94	86	90
Lagos	82	82	86
Kaduna	83	70	70
Enugu	87	56	85
Jigawa	95	80	95
Kano	97	82	88
Notes: Figures in this table reflect percentage of schools/SBMCs which meet the logframe standard. This table only contains Output 4 phase 1 pilot and phase 2 schools which have reached the mentoring/monitoring stage of SBMC development. All data comes from SMO reporting of SBMCs latest as of July 2014			

Note. + / - indicate a statistically significant positive/negative difference ($p < .05$); blank cells indicate no statistically significant difference.

SBMC functionality

The SMO report notes that the functionality scores are generally encouraging and scores of the second phase roll-out schools (not disaggregated in Table 15) also indicate that the newer SBMCs are catching up quickly with phase-1 pilot schools on all indicators of SBMC functionality. The SMO report notes that this may be due to there being more general capacity at state, LGEA and local level to support SBMC development and that information about SBMCs has reached new schools from the pilot ones (referred to as a 'multiplier effect' in the report). CS2 results on SBMC functionality, discussed in section 5.1 above, found that around 60% of ESSPIN schools are meeting the functionality logframe standard compared to 12–13% of the control schools.

The SMO report also notes the improved record-keeping being practised by SBMCs as a result of ESSPIN training. Learning from the implementation of CS1 in 2012, ESSPIN's Community Engagement team provided additional technical support to SBMCs on tracking SBMC development. This meant ensuring that SBMCs were able to provide appropriate evidence to SMOs of progress, including SBMC meeting minutes, financial records, school visitor books, receipts of purchases, SDPs etc.

The SMO report also notes the impact of ongoing insecurity in Kano State on SBMC functioning. The reports highlight that SBMCs are adjusting meeting times to work around the insecurity and that there is significant anxiety in and around schools about safety and security. It is possible that this will be reflected in the Kano results in future.

5.4.1 Women's and children's inclusiveness within SBMCs

The SMO report cites formative SBMC research (Poulsen, 2009) which has highlighted that women's participation in some states and in some communities within states was 'highly constrained' and that the participation of children was 'not accepted' due to sociocultural norms. So whilst the women's and children's inclusiveness results look weaker when compared to the other two indicators, it is important to see them in that context. The SMO report highlights the case of Kaduna in particular:

> *SBMC research in 2009 highlighted Kaduna State to be one of the states where women and children's participation was lower due to sociocultural and religious norms and that it was not common in many areas of men and women to sit together and discuss issues or to include children in adult discussions.* (Kaduna State SBMC report)

According to the SMO report, work to support women and children, including sustained capacity development and sensitisation, appears to have gradually brought about change from this situation, but these are challenging areas in several states, particularly the northern ones, due to sociocultural and religious norms. Overall, CS2 results (section 5.2, 5.3) found that 2–3% of non-ESSPIN schools meet the women's inclusiveness criteria compared to 44–48% ESSPIN schools (section 5.2), and 2% (control) vis-à-vis 15–21% (ESSPIN) for children's inclusiveness (section 5.3).

5.4.2 Communities supporting inclusive education

The SMO report states that the overall results indicate 'a very high number' of SBMCs active in mobilising resources for their schools from a wide range of community sources, and that they are using these to support access for marginalised children and to improve schools. Some of this is also corroborated by the CS2 results discussed above (see Table 10 for instance): for example, ESSPIN schools were twice as likely to meet the logframe criterion for 'raised issue of children's exclusion' than the control schools.

The SMO report also mentions that despite the newness of the inclusive education concept in school communities, there has been broad acceptance and support for all children to attend, remain in and learn better in an improved and safer school environment. Qualitative information from SMO reports highlights that SBMCs are supporting children from nomadic communities to attend and stay in school (Kaduna, Kwara and Kano), girls in some LGEAs (Kano), children from poorer backgrounds and children with a disability (with Kwara and Enugu giving specific attention to disabled children going back to mainstream schools), and particularly taking action around child protection issues which arise in and around the school (Kaduna).[15]

The SMO report raises issues linked to the language used in the classroom as a major barrier for some children, especially children from nomadic communities who speak Fulfulde, and do not understand Hausa or Yoruba. These children are dropping out of school quickly even if they enrol. They also highlight higher numbers of children affected by a disability now attending mainstream schools and advocate more assistance for teachers and schools to fully include these children. Issues of disability (section 7.7) and minority language (section 7.5) are discussed below in more detail.

5.4.3 Estimated enrolment and return of children to school as a result of SBMC action

In 2013, SMO reporting formats were adapted to capture SBMC estimates of numbers of pupils enrolled in school, and numbers of pupils who had dropped out of school as a result of their actions (actions such as providing pupils from poorer families with uniforms/text books/learning materials, supporting access and learning for pupils affected by a disability, conducting house-to-house visits, and engaging with nomadic communities etc.). The SMO report notes that to date it has been possible to get realistic estimates only from Kwara State, where the SMOs reported that between 2011 and 2014, approximately 20,000 pupils were enrolled in school due to SBMC action and 3,000 pupils were successfully enrolled back in school due to SBMC action.

[15] States named in parentheses above are those mentioned specifically in the SMO report as taking action, but this does not mean that other states are not also taking action.

PART B: DIFFERENCES IN EDUCATION OUTCOMES BY GENDER AND BACKGROUND

6 Introduction to Part B

Do boys in school learn more than girls? Are students from poorer families, overage students, and speakers of minority languages learning less than their peers? Are female teachers and head teachers more competent or effective than their male colleagues? And are any such gaps in learning or competence larger in ESSPIN schools than in other schools?

Part B attempts to answer these questions by focusing on differences in pupil learning outcomes by gender, wealth, location, language, age, and disability; and on gender differences in teacher competence and head teacher effectiveness. Our key indicators are the results of pupil tests in literacy and numeracy at grade 2 and 4, and standards of teacher competence and head teacher effectiveness based on interview responses and lesson observations (see section 2.2 above). In each case we disaggregate by the relevant dimension (gender, wealth, etc.) and apply statistical tests to assess whether any differences in our estimates are likely to have occurred by chance or to represent real differences in the population.

Where relevant and feasible, we also disaggregate further by whether or not the school received ESSPIN's set of interventions aiming at strengthening the capability of primary schools to provide improved learning outcomes. 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 head teachers. 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 help maintain and continue gains in quality.[16] This disaggregation gives us an indication of whether ESSPIN school improvement interventions are reaching all groups of pupils, teachers and head teachers equally, or benefiting some more than others.

For ease of reading in the following sections we refer to schools that have received **at least one year of full package of ESSPIN Output 3 intervention prior to 2013–14** as ESSPIN schools, and other schools as non-ESSPIN schools.[17]

[16] Scaling-up of the school improvement programme is discussed in more detail in the overall CS2 report (Cameron, 2015).
[17] Thus, the non-ESSPIN schools include some which received intervention in 2013–14. This intervention is unlikely to have had an effect by the time of the survey, which was towards the end of the same school year, and so it is more appropriate to class these schools with the non-ESSPIN group than the ESSPIN group.

7 Pupil learning outcomes

In this section, we discuss differences in pupils' learning outcomes by gender, household wealth, remoteness of school, language, age and disability. Pupils in both rounds of the survey were administered grade 2 and 4 literacy and numeracy tests. We find no significant gender differences. Pupils from the richest 20% of households answered twice as many questions correctly as those from the poorest. We find evidence that the poorest children are benefiting disproportionately from ESSPIN's Output 3 interventions, so that treatment schools have smaller wealth disparities than control ones. On the remoteness of schools, we find that pupils in urban schools performed significantly better than those in rural schools. The rural–urban gap, though less pronounced in ESSPIN schools, is still significant. We do not find any evidence that pupils who predominantly speak a minority language (i.e. not Igbo in Enugu; Hausa in Jigawa, Kano and Kaduna; and Yoruba in Kwara) received significantly different marks. Across tests and grades, overage pupils performed better than their counterparts who were age-appropriate for their grades. The CS2 pupil tests included a number of questions on assessing various forms of physical disability among sampled pupils; in total, 54 pupils (~0.5%) from the CS2 sample were found to be disabled. As these children cannot be taken as a representative sample—and were disabled in different ways—we have not analysed their test results separately.

7.1 Enrolment of boys and girls

Before discussing pupil learning outcomes, we briefly present and discuss descriptive statistics in Table 16 on the enrolment of girls and boys in primary grades of government schools between 2009 and 2013 obtained from the ASC.

The ratio of girls to boys in primary grades is farthest from 1 (i.e. fewer girls than boys) in Jigawa and Kaduna, and closest to equality in Kano, Lagos and Enugu. However, both Jigawa and Kaduna seem to be moving gradually towards greater equality. The southern states (Enugu, Kwara, and Lagos) have the highest proportion of schools with equal numbers of girls and boys in primary grades (though Kaduna's figures seem comparable to Kwara at least). Jigawa, Kaduna and Kano have all progressed more than the other states in coming closer to equal numbers of boys and girls.

Table 16: Enrolment of girls and boys in primary grades of government schools, 2009–13

	Ratio of girls to boys in primary grades		% schools with equal numbers of girls and boys (+/- 5%)		Parity change
	2009	2013	2009	2013	
Enugu	0.98	0.95	21	18	-0.03
Jigawa	0.70	0.76	5	7	0.06
Kaduna	0.81	0.86	14	17	0.05
Kano	0.90	0.97	6	9	0.07
Kwara	0.89	0.91	14	15	0.02
Lagos	1.04	1.04	28	26	0.00

Note. + / - indicate a statistically significant positive/negative difference ($p < .05$); blank cells indicate no statistically significant difference.

7.2 Gender differences

Although there appear to be more boys than girls enrolling in some states, the results from CS2 suggest that, once in school, boys and girls obtain roughly equal learning outcomes (Table 17).[18] There was no significant difference between boys and girls in average test scores in literacy and numeracy tests in grades 2 and 4. This largely corresponds with gender-disaggregated pupil test scores from CS1, where scores on N2, L2, and L4 did not differ significantly along the gender dimension, though girls scored significantly better (42%) than boys (38%) on the grade 4 numeracy test in CS1.

Table 17: Gender disaggregated pupil learning outcomes in CS2

Mean test score %	Boy	Girl	Significant diff.
N2	38.2	37.3	
L2	29.9	30.3	
N4	33	32.9	
L4	28.9	30.2	

Note. + / - indicate a statistically significant positive/negative difference (p < .05); blank cells indicate no statistically significant difference.

Did girls in ESSPIN schools perform differently from boys on the tests? We test this by looking at the results by gender and intervention group (Table 18). There is no evidence of gender differences in either ESSPIN or non-ESSPIN schools. We also disaggregate gender-based pupil test scores by states (not shown here) and find no significant differences in mean test scores, except for Lagos, where girls (67.8%) perform better than boys (61.1%) on the grade 4 literacy test (L4).

Table 18: Gender-disaggregated pupil learning outcomes in CS2 by intervention groups

Mean test score %	Non-ESSPIN		Sig. diff.	ESSPIN schools		Sig. diff.
	Boys	Girls		Boys	Girls	
N2	34.6	33.8		47.9	46.2	
L2	26.2	26		40.4	41.3	
N4	30.5	29.5		38.7	40.2	
L4	24.7	24.7		38.4	42	

Note. + / - indicate a statistically significant positive/negative difference (p < .05); blank cells indicate no statistically significant difference.

7.3 Wealth

Did poorer pupils perform differently from pupils from wealthier backgrounds on the CS2 pupil tests? In CS2, data on pupil household wealth was collected using a short asset questionnaire administered at the end of the grade 4 literacy and numeracy tests. The data collectors showed grade 4 pupils pictures of common household assets (given in Box 5), asking the pupil if his/her family had that asset at home.

[18] The differences in enrolment probably mean that the sample of girls in school is more selective than the sample of boys (i.e. the girls are a more able group because less advantaged girls do not attend) so that gender parity in test scores does not imply that schools are equally effective in educating girls and boys.

Box 5: Question in CS2 on pupil household wealth

I just want you to look at some pictures of things. Please look at these pictures one by one and tell me if your family has any of these things in your home.

Interviewer: point to first picture.

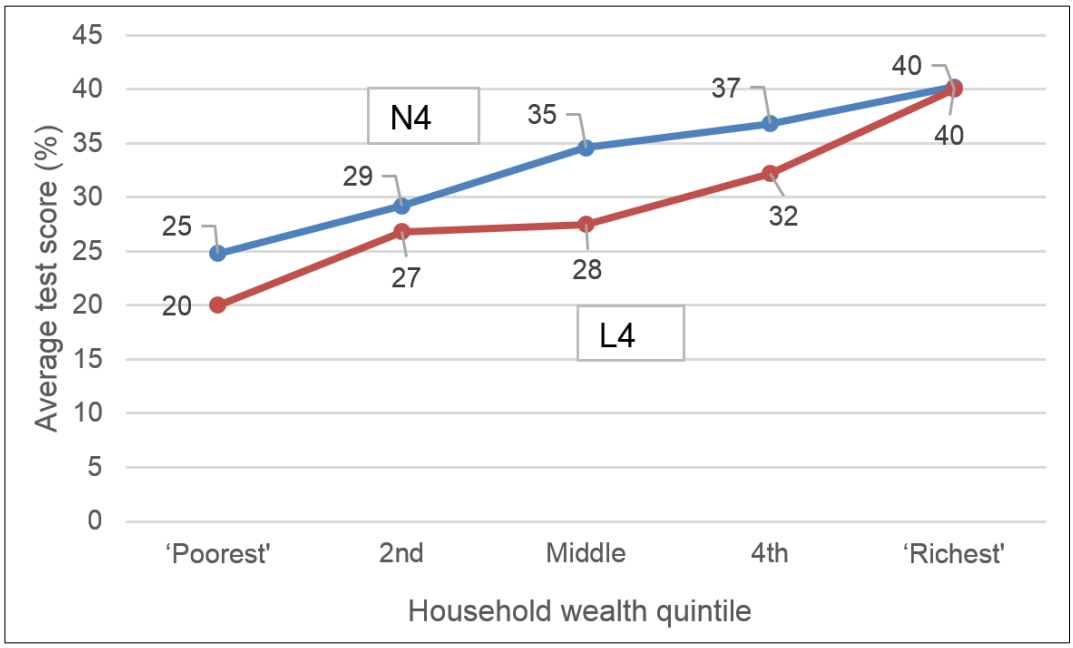
Does your family have a sofa?

Repeat for subsequent items.

Sofa, chair, table, mattress/bed, mat, sewing machine, fridge, bicycle, motorbike, car, Generator, fan, radio, TV, computer, mobile phone.

A household asset index and wealth quintiles were created on this basis.[19] It should be noted that the wealth groups are calculated relative to other pupils in the same state, i.e. 'poorest' means the poorest 20% in Lagos, plus the poorest 20% in Enugu, plus the poorest 20% in Kwara, etc., rather than the poorest 20% in the six states combined.[20] This indicator strongly predicts test scores (Figure 4). On average, pupils from the richest 20% of households are answering almost twice as many questions correctly as those from the poorest, and the correlation between test scores and household assets is more pronounced for literacy than for numeracy.

Figure 4: Average scores in grade 4 numeracy and literacy tests for all CS2 schools by household asset quintile



Does being in an ESSPIN school in CS2 dampen the effect of inequality in household assets? Average learning outcomes for the poorest are under 20% in non-ESSPIN schools but around 30% in ESSPIN schools (the blue bars in Figure 5). For the richest (the red bars in Figure 5), average test scores are 39% in non-ESSPIN schools but vary from 41 to 48% in ESSPIN schools. This

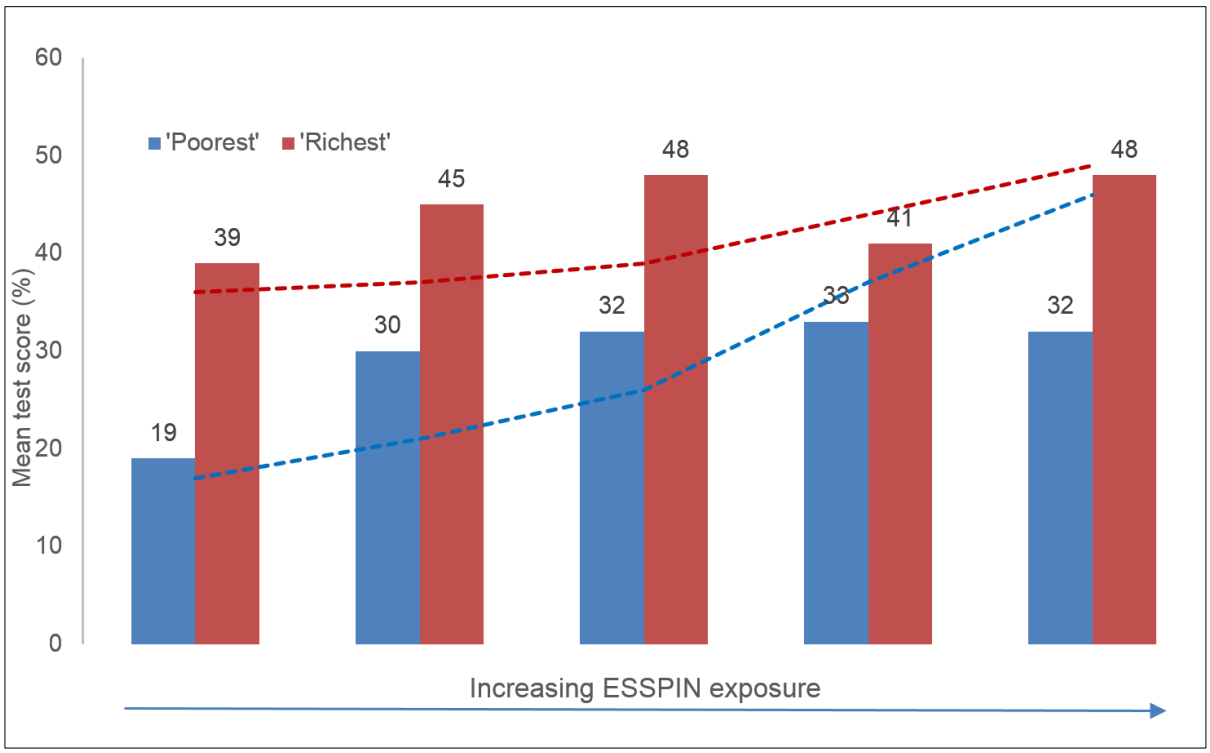
[19] The index was calculated by taking the first component in the principal components analysis (as in Filmer and Pritchett, 2001).

[20] Such within-state comparison of wealth differences is not intended to suppress wealth variations between states, which is itself an axis of marginalisation in that pupils in the most disadvantaged states could be considered the most excluded, rather more than sub-groups across the whole sample. In other words, pupils from wealthier families in disadvantaged states may not necessarily fare better than poorer pupils in more advantaged states on test scores.

results in a wealth gap—a difference between test scores for the richest and poorest—of 20 percentage points in non-ESSPIN schools, and between 8 to 16 percentage points in schools receiving different levels of ESSPIN exposure.

We confirm that this is a significant difference in wealth gap using a statistical model,[21] which we also use to plot the average relationship between wealth, ESSPIN intervention and test scores—the dotted trend lines in Figure 5 (smoothing out some inconsistencies in effect between different levels of ESSPIN intervention to produce an overall trend). Learning outcomes appear to be higher for both rich and poor students in ESSPIN schools compared to non-ESSPIN, but more for the poor students, so that the wealth gap is smaller in ESSPIN schools. However, we are unable to establish causation, because the selection of ESSPIN schools is not random and we do not have data on the wealth of children at different points in time. It is possible that children in ESSPIN schools were more equal at the beginning of the intervention, for example because they are more likely to be in urban or peri-urban areas, where richer and poorer children mix within the same classrooms. Nevertheless, the convergence of rich and poor students in schools with more ESSPIN intervention is at least consistent with a reduction by ESSPIN in the effects of inequality in socioeconomic status on learning outcomes.

Figure 5: Does being in an ESSPIN school reduce the effect of wealth inequality?



Note: Bar graphs show actual values, while dotted lines show average trend in mean scores.

When we disaggregate grade 4 literacy and numeracy scores by states, we do not find any such effect of being in an ESSPIN school at the state-level; the reason for the lack of a similar result here is not entirely clear and presumably, at least partly, is due to reduced sample sizes at the state level vis-à-vis aggregate sample size across all six states.

[21] The model is a linear regression of average test score on grade 4 literacy and numeracy, an indicator of wealth relative to other children in the same state, an indicator of the extent to which the child's school has been exposed to ESSPIN, and an interaction term between these two variables. A significant negative coefficient on the interaction term indicates that wealth inequality in learning outcomes is smaller in schools with more ESSPIN intervention.

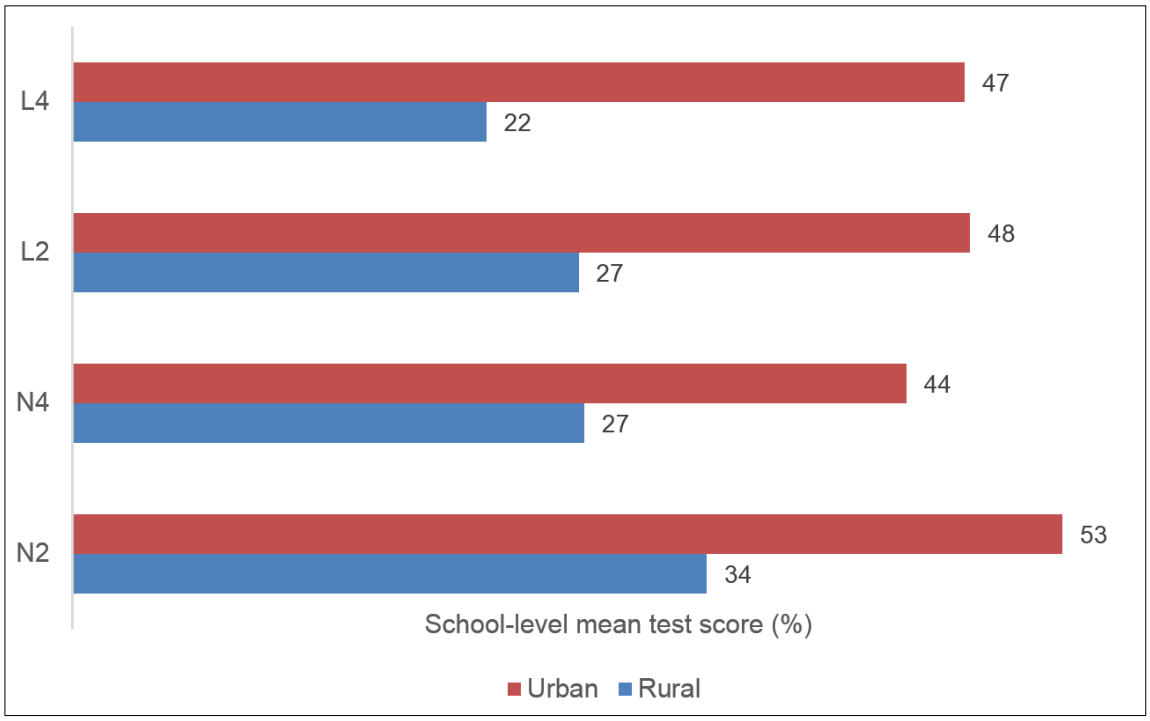
7.4 School location

Do pupils in schools which are farther away from urban areas, the state capital or local government area (LGA) headquarters have different learning outcomes compared to schools which are less remote?

7.4.1 Rural schools

We supplemented the CS2 data with information from the ASC on whether a school was 'urban' or 'rural' and used this classification to disaggregate average pupil test scores per school for each of the four tests (Figure 6). Within all CS2 schools, we find that pupils in urban schools performed significantly better than those in rural schools by factors of 1.5–2, depending on the test. The magnitude of this rural–urban gap is much more pronounced for literacy test scores and more so for grade 4 than grade 2. Urban pupils in grade 4 did twice as well as their rural counterparts on the literacy test.

Figure 6: CS2 Pupil test scores in rural and urban schools



Is this rural–urban gap diminished in ESSPIN schools? There are large rural–urban gaps in both ESSPIN and non-ESSPIN schools (Table 19). For numeracy scores, this gap (in percentage point terms) is greater in the control schools, while for literacy, it is greater in ESSPIN schools. This suggests that ESSPIN may be associated with better outcomes in both rural and urban schools, but has so far had little impact on the size of the gaps between them. Again, however, we are not able to establish causation (or a lack of it), and this pattern may reflect the way that ESSPIN schools were before the intervention rather than what has happened since.

Table 19: CS2 pupil test scores in rural and urban schools, by ESSPIN status

	Control school				Treatment school			
School-level mean (%)	Rural	Urban	Significant difference	Diff. (pp.)	Rural	Urban	Significant difference	Diff. (pp.)
N2	31.3	47.7	+	16.4	43.5	58.1	+	14.6
N4	25.5	42.1	+	16.6	33.3	46.5	+	13.2
L2	24.7	39.5	+	14.8	35.7	57.2	+	21.5
L4	20	39.3	+	14.6	29.6	55.7	+	26.1

Note. + / - indicate a statistically significant positive/negative difference (p < .05); blank cells indicate no statistically significant difference.

7.4.2 Distance of school from LGA headquarters

Schools that are more remote typically have lower test scores. We estimate that for every 10 kilometre increase in distance between the school and the LGA headquarters, test scores drop by between 0.7 and 1.3 percentage points (Table 20).[22] The drop is statistically significant for L2, L4 and N4. The remoteness effect appears to be strongest in Enugu, Jigawa and Kano, and overall seems more pronounced for English literacy scores than numeracy. Remoteness here appears mainly to be highlighting differences between rural and urban areas, and so confirms the findings in the previous section. When we apply the same analysis to rural areas only (the results are not shown here separately), the effects of distance from LGA headquarters are no longer significant.

Table 20: Change in test scores per 10-kilometre increase in distance from school to LGA headquarters

Test scores (pp.)	L2	L4	N2	N4
Enugu	-2*	-2.3*	0.7	-0.4
Jigawa	-1.6*	-2*	-3*	-2.6*
Kaduna	-0.8	-0.2	-0.5	-0.3
Kano	-2	-5.0*	-2.9*	-3.4*
Lagos	-1*	-1.4*	0.6*	-0.5
Total	-1*	-1.3*	-0.7	-0.9*

Note. * indicates a statistically significant change (p < .05).

7.5 Speaking a minority language

In CS2, sampled pupils in both grades 2 and 4 were asked to name the main language they speak at home. Box 6 shows the question asked in the pupil tests and the options provided.

Box 6: Question in CS2 on main language spoken at home by pupils

What language do you mostly speak at home?

Answer options: Benin / Edo; Ebira; English; Esan; Fulfulde; Hausa; Gbagyi; Gbari; Ibibio; Igbo; Kanuri; Nupe; Yoruba; Other; Don't know; Refused to answer.

As would be expected, most children said they speak Hausa in Jigawa, Kaduna and Kano; Igbo in Enugu; and Yoruba in Kwara and Lagos (Table 21). Lagos is more mixed than the other states with

[22] We use a linear regression of test score on distance from LGA headquarters, using information on the latter obtained from the Nigerian School Census 2013–14: National Bureau of Statistics (2013).

a greater proportion of children who speak English or Igbo at home. Overall 16% of pupils in the five states excluding Lagos speak a minority language at home.

Table 21: Main language spoken by pupils at home in CS2, by state

Language (%)	Enugu	Jigawa	Kaduna	Kano	Kwara	Lagos
Benin / Edo	0	0	0	0	0	2
Ebira	0	0	0	0	0	1
English	6	0	2	0	2	23
Esan	0	0	0	0	0	0
Fulfulde	0	5	2	5	2	0
Hausa	0	93	63	94	3	4
Gbagyi	0	0	4	0	0	0
Gbari	0	0	0	0	0	0
Ibibio	0	0	0	0	0	1
Igbo	92	0	0	0	1	9
Kanuri	0	2	0	0	0	0
Nupe	0	0	0	0	7	0
Yoruba	0	0	1	0	78	54
Other/missing	2	0	28	0	7	8
Total	100	100	100	100	100	100

Note. The numbers in red indicate the main language of each state.

Do pupils who predominantly speak a minority language have different learning outcomes than those who speak the main language of a state? We do not find any evidence of pupils who predominantly speak a minority language scoring significantly different marks in the numeracy and literacy tests, compared to pupils who speak the state's majority language at home.[23] This is true at individual state levels and for five states together(Table 22).[24] We do not find any difference between minority and majority language speakers within either ESSPIN or non-ESSPIN schools (Table 23).

Table 22: Pupil test scores by language status in CS2

Mean test scores (%)	Majority	Minority	Significant difference
N2	35.9	35.2	
L2	26.7	30.7	
N4	31	32.2	
L4	25.8	26.7	

[23] The numeracy assessments were administered in Hausa. A number of the English literacy items also used Hausa, except those where comprehension of the question in English was essential to the competency being tested by the item.
[24] We exclude Lagos from this analysis because it appears to have a more diverse language situation than other states, with little more than half of the children saying that they speak the majority language (Yoruba) at home. Children in Lagos are likely to mix languages at home and in their daily lives, both within and outside of school. In each of the other states one language more clearly dominates (although some degree of switching between the state's main language and English is common.)

Table 23: Pupil test scores in CS2, by language status and Output 3 intervention group

%	Grade 2	Grade 4	Total
Overage	38	28	34
Distribution of overage pupils across states (%)			
Enugu	4	4	4
Jigawa	6	10	7
Kaduna	17	16	16
Kano	63	59	62
Kwara	3	4	3
Lagos	7	7	8
Total	100	100	100

Note: 11,015 pupils were tested across grades and tests in CS2. Age data was missing for 6%, 4% and 4% pupils in class 2, 4 and overall.

Thus, speaking a minority language at home does not, on the basis of this evidence, appear to be a major impediment to learning. The underlying reasons for this are not entirely clear, raising some additional questions for research to test possible hypotheses:

- **Do pupils speaking minority languages come from 'different' backgrounds, say, in terms of household characteristics?** We cannot easily test whether minority-language-speaking pupils are any different from the majority in terms of household characteristics. Their numbers in the pupil sample are relatively small to test this statistically with reliability. We also hypothesise that children who do not speak the majority language at all are much more likely not to be in school, and so would not have participated in our survey.

- **Does the national education policy on switching from mother-tongue instruction to English in grade 4 confer a degree of language disadvantage for all pupils?** It is not clear that this language policy, in practice, is a disadvantage for all students uniformly, accounting for other background factors, given the widespread use of English across grades despite this policy. Moreover, if this were the case, we would have seen the emergence of language disadvantage in earlier years (when instruction is in the mother tongue) but gradually less so in higher grades of primary school.

7.6 Being overage for one's grade

What does CS2 say about learning outcomes among pupils who are overage for their grades? To investigate this question, within each grade, i.e. within grade 2 and grade 4, we looked at each pupil's test score in relation to whether or not is the pupil was overage. Pupils are expected to be seven or eight years old in grade 2 and nine or 10 years old in grade 4. We define overage as children who are nine years or older in grade 2 or 11 years or older in grade 4.

7.6.1 Profiles of overage pupils in CS2

Here we discuss their numbers and profiles in terms of which states they come from, their household wealth and the extent to which they are overage, compared to their appropriately aged peers. Table 24 below shows that around 34% of the pupil sample from CS2 (i.e. 38% of the grade 2 and 28% of the grade 4) are overage, and they come mainly from Kano (62%) and Kaduna (16%).

Table 24: Proportion of overage pupils in CS2 overall and by state

%	Grade 2	Grade 4	Total
Overage	38	28	34
Distribution of overage pupils across states (%)			
Enugu	4	4	4
Jigawa	6	10	7
Kaduna	17	16	16
Kano	63	59	62
Kwara	3	4	3
Lagos	7	7	8
Total	100	100	100
Note: 11,015 pupils were tested across grades and tests in CS2. Age data was missing for 6%, 4% and 4% pupils in grade 2, 4 and overall.			

Furthermore, Table 25 below, on the extent of overage, shows that in both grades 2 and 4 a large proportion of overage pupils are older than their appropriately aged peers by one year. We also find that overage pupils in the CS2 sample are from significantly 'wealthier' households compared to their appropriately aged peers (not tabulated separately).

Table 25: Extent of overage among CS2 pupils (% of overage pupils)

Older by	Grade 2	Grade 4
One year	59	62
Two years	17	23
Three years	15	9
Four or more years	8	6
Total overage (%)	100	100

7.6.2 Results from CS2

How do overage pupils perform compared to the rest of the pupil sample? Table 26 and Table 27 below show pupils' test scores by the degree to which they are overage. Across tests and grades, overage pupils in CS2 performed better than their counterparts who were age-appropriate for their grades. Focusing on ESSPIN schools alone, the results are largely intact across both ESSPIN and non-ESSPIN schools: overage pupils are performing significantly better across tests and grades.

Table 26: Mean scores across tests in all CS2 schools by overage status

Mean test score (%)	Not overage	Overage	Significant difference
N2	34.4	48.4	+
N4	29.9	39.1	+
L2	27.1	36.3	+
L4	26.6	35.5	+

Note. + / - indicate a statistically significant positive/negative difference ($p < .05$); blank cells indicate no statistically significant difference.

Table 27: Mean scores across tests in all CS2 schools by overage status and intervention group

	Non-ESSPIN			ESSPIN schools		
Mean test score (%)	Not overage	Overage	Significant difference	Not overage	Overage	Significant difference
N2	29.8	46.4	+	45.3	53.6	+
N4	28.2	37.7	+	39.0	43.3	+
L2	22.4	33.2	+	38.9	44.9	+
L4	22.6	32.8	+	39.8	48.6	+

Note. + / - indicate a statistically significant positive/negative difference ($p < .05$); blank cells indicate no statistically significant difference.

7.6.3 Does being 'overage' for one's grade really matter?

There is wide variation in pupils' ages within grades in many developing countries, which means that pupils of different ages, even though they face unique age-specific learning challenges, are enrolled in the same grade and thus are exposed to the same learning material. A number of studies have looked at the association between various indicators of interest pertaining to attainment and achievement vis-à-vis the pupil's age in the context of developing countries. We discuss these below, with the caveat that only limited extrapolation is possible to the Nigerian context. This is due to two reasons: (1) Nigeria has one of the lowest official grade repetition rates (3%) among Anglophone sub-Saharan African countries (Ndaruhutse, 2008);[25] and (2) disaggregation of CS2 pupil household wealth index data by age, as mentioned above, shows that overage pupils were from 'wealthier' households compared to their appropriately aged peers.

The Education Policy and Data Centre (EPDC) (2011) looked at promotion, repetition, and dropout rates for overage, appropriately aged and underage pupils using the demographic and household survey data from 35 low- and middle-income countries. It found that in the early grades of primary school, promotion rates are positively correlated with age; the overage students generally have the highest promotion rates. In the later grades of primary school, however, the study found that overage students encounter growing pressure to drop out. The promotion and survival rates of overage pupils in these grades are lower and, in most countries, the dropout rates are higher than those of appropriately aged or underage pupils. The dropout rates for overage students gradually increase through primary school and in many countries peak in the last grade of primary. It is not known to what extent this effect is driven by external demands and opportunities for older children and to what extent it is driven by factors internal to the school environment, such as the effect of being overage on a student's own learning trajectory. However, as we mention above, the Nigerian context appears different from the 35 countries studied by EPDC, due to the low grade repetition rates in Nigeria. In terms of the association between overage pupils and household background, Lewin and Sabates (2011) looked at trends in patterns of enrolment by age for 13 African countries while Taylor et al. (2010) looked at South Africa in particular. Both studies found that overage children were more likely to come from poorer homes than their appropriately aged peers, in contrast to the CS2 pupil sample. The South African study also found that overage children performed less well on tests than their appropriately aged peers and gained less learning from one year to the next as they progressed from grades three to five.

Lloyd (2012) reviews the literature (including some of the studies cited above) on classroom age distribution and concludes that existing studies provide evidence that overage students are more likely to drop out as primary school progresses because their learning increasingly falls behind. Among those overage students who remain enrolled, their prospects appear increasingly poor,

[25] The highest grade repetition rates were reported in Lesotho (21%), Malawi (16%) and Swaziland (15%).

decreasing the likelihood that they will score sufficiently well in their primary school leaving exams to be eligible for a place at secondary school. In early grades, however, being overage for one's grade may not be a disadvantage, particularly for literacy acquisition.

Perhaps the key pupil characteristics is being 'ready' for the instruction provided—curricula are often deemed to be over-ambitious for the grade concerned (Pritchett and Beatty, 2014) and if overage pupils have received more preparation through private pre-schooling, this may explain their better performance. Parents who are able to afford pre-schooling may enrol their children in school when they are 'ready', i.e. adequately prepared, whereas poorer pupils may enrol as soon as their families are able to, their parents not being able to afford other forms of pre-schooling.

Ultimately, it is unclear from the literature cited above (or from a school survey such as CS2, which does not collect extensive household-level data) whether these differences between past studies and the current study reflect underlying household choices pertaining to education, particularly age at enrolment, exposure to pre-schooling or sporadic drop-out and return. This uncertainty is an area for further research. Our findings that overage pupils in the CS2 sample have significantly higher test scores and come from better-off households indicate interesting differences in the classroom age and learning profiles in Nigeria, which, as mentioned above, is also reported to have one of the lowest grade repetition rates among Anglophone sub-Saharan African countries.

7.7 Disability

After discussions with ESSPIN and external disability specialists, the CS2 pupil tests included a number of questions on assessing various forms of physical disability among the pupils sampled. These questions were added at the beginning of each test. In each case, children who do not have a particular ability were not made to sit through questions that required that ability, e.g. children with impaired vision were not asked to read a text from the book.

Whereas in CS1 the question of disability was not explicitly addressed, for CS2 it was made clear to data collectors that children with disabilities needed to be included in the test, but that where they did not possess particular abilities for answering specific questions, the CAPI program would skip these questions. The data collector manual included careful instructions for data collectors to use common sense when applying the ability test questions. For example, a child who responded when their name was called but did not respond in any way to a greeting would not be dismissed as 'unable to hear', but as shy or unable to understand the greeting.

However, it should be noted that none of the four test used in CS2 aimed to assess disability—which would require a much larger set of specific questions with follow-up by specialists. The focus instead was on a pupil's ability to take the test. Thus, these were not intended as proper tests of disability, but as an indication of whether sampled pupils possessed the abilities to see, hear, speak and write needed for different test items.

7.7.1 Ability to hear

All of the questions in both literacy and numeracy tests require the pupil to understand spoken instructions. For children who were unable to hear or understand through some other means such as lip-reading or with the aid of a signing assistant, the CAPI software therefore skipped to the end of the test.[26] Data collectors were trained in ascertaining hearing impairment using the following instructions (Box 7).

[26] For pupils with hearing impairment, provision of written instructions only were considered an inappropriate substitute as all questions required verbal instructions and explanation from the data collector.

Box 7: Instructions to data collectors for screening hearing-impaired pupils

Interviewer: While leading the pupil to the test location, greet the child and ask his/her name, using the local language where possible. When you first speak to the pupil, if the pupil shows no signs of hearing what you say, confirm with the teacher whether the pupil can hear. If he or she cannot hear but can understand through some other means—lip-reading or through a signing assistant—continue with the test. If the pupil is not able to understand you, give them the biscuit, drink and pencil, and thank them for their participation.

7.7.2 Ability to speak

Pupils who appeared unable to speak at all were not given questions that required a spoken answer. In the numeracy tests, their ability to speak was checked in the following way.

Box 8: Instructions to data collectors for screening pupils unable to speak (numeracy test)

Prompt if necessary until you get a reply to your greeting.

Mark whether the pupil responds verbally to your greeting and/or to saying his name.

In the literacy tests, this was combined with a question checking whether the child could respond in English.

Box 9: Instructions to data collectors for screening pupils unable to speak (literacy test)

Greet the child again in English:

Good morning / good afternoon

Prompt if necessary until you get a reply to your greeting.

Mark whether the child responds with 'Good morning/afternoon' or any culturally appropriate greeting; responds verbally but inappropriately; or does not respond at all.

If the pupil responded verbally but inappropriately, they were marked incorrect but the test continued without skipping any questions. If they did not respond at all, they were marked as unable to speak, and spoken questions in the test subsequently skipped.

7.7.3 Ability to see

Children were asked the following question to gauge whether they could see well enough to take the test. If they could not then they were subsequently asked only questions that could be asked orally and required an oral response.

Box 10: Instructions to data collectors for screening visually impaired pupils

I am going to ask you lots of number questions. I will ask you to write or say the answers. You should try your best but do not worry if you cannot answer.

Can you see the book here?

Point to the pupil book on the table.

Mark whether the pupil indicates that he/she can see the book, by looking at it and/or saying yes.

7.7.4 Ability to write

The following question tested whether children had the physical ability to hold a pencil and mark the page. For children who could not do this, questions requiring writing were skipped.

Box 11: Instructions to data collectors for screening pupils unable to speak (numeracy test)

Interviewer: Turn to the 'Drawing' page and give the child the pencil.

I'm going to draw a line between these two dots.

Interviewer: draw a line on the pupil book between the two dots at the top of the page.

Now, can you do the same and draw a line between these two dots?

Interviewer: point to the two dots lower down the page.

Mark whether the pupil draws or writes something on the book, regardless of whether it is a straight line between the two dots or something else.

7.7.5 Descriptive statistics from CS2

In total, 54 pupils from the CS2 sample (i.e. less than 0.5%) were unable to perform one or more of the tasks as per the four screening questions (Table 28). As these children cannot be taken as a representative sample—and were disabled in different ways—we have not analysed their test results separately.

Table 28: Descriptive statistics on disabled pupils from CS2

	N2	L2	N4	L4	Total
All respondents	2,784	2,774	2,725	2,733	11,016
Pupils with disabilities					
Unable to hear	0	5	1	4	10
Unable to speak	5	3	1	2	11
Unable to see	3	4	1	3	11
Unable to write	8	9	1	4	22
Total	16	21	4	13	54

7.7.6 An 'ideal' test of disability?

In general, monitoring progress in education among individuals, and particularly children, with disabilities is challenging and has not been possible in most countries. A recent blog post (Mont, 2014) notes, 'In fact, until recently there were no generally agreed upon questions for identifying people with disabilities that had been tested widely in developing countries. Fortunately this situation is changing.'

Recently the UN Statistical Commission's Washington Group on Disability Statistics (WG) has developed and tested a short set of six questions for identifying people with disabilities, with the aim of including these as a regular part of every national census surveys, including demographic surveys, household and expenditure surveys, and labour market surveys. However, these questions are not suitable for children and so the WG and UNICEF have recently piloted a set of child disability questions in several countries; this is near finalisation and they are likely to be

included in the UNICEF Multiple Indicator Cluster Surveys. Even with these child disability questions, in school surveys such as the CS2, looking only at individual children and not at households misses the broader impact of disability on children's lives in general, and how this might impede their interaction with the education system. Nevertheless, it would be a major step forward.

As is apparent from the discussion above, monitoring the extent to which children with disabilities are attending and excelling at school is complex and still an emerging area within education metrics. As mentioned above, none of the pupil tests in CS2 aimed at assessing disability; their focus was rather on a pupil's ability to take the test. The tests were not designed or adapted to cater to pupils with one or more physical or intellectual impairment. Thus, these were not intended as proper tests of disability, but as an indication of whether the sampled pupils possessed the abilities needed for different test items.

8 Teacher competence: gender differences

In this section, we present teacher competence results disaggregated by gender. Overall, 50% of CS2 teachers were female, with wide variation between northern and southern states. Female teachers performed significantly better than their male counterparts on almost all the logframe teacher competence criteria. Part of the difference in competence may be due to most female teachers being located in states where teachers are, in general, more competent. But even within states, there are still large differences in the proportions of female and male teachers reaching competence standards. Female teachers were also more likely to encourage participation of children sitting in different parts of the classroom, although there was no significant difference between male and female teachers in terms of the equal participation of boys and girls in the lesson.

8.1 Gender composition of CS2 teacher sample

Overall, half of the teachers across the six states are female (Table 29). There is wide variation between the states. In the north, Jigawa had 14% female teachers, with Kano and Kaduna at 24 and 56 % respectively. Lagos and Enugu had the highest proportions of female teachers (84%). Most female teachers in the six states are therefore based in Lagos.

Table 29: Percentage of public primary school teachers who are female

State	Female teachers (%)
Jigawa	14%
Kaduna	56%
Kano	24%
Enugu	84%
Kwara	73%
Lagos	84%
Total	50%

Note. The proportions shown are based on the CS2 sample with weights applied, so should be representative of the population. Teachers who teach only religion or who did not teach grades 1 to 6 were excluded.

8.2 Teacher competence criteria

The ESSPIN logframe sets four criteria for judging the competence of teachers (Box 12). Female teachers appear to be performing significantly better than their male counterparts on almost all the logframe teacher competence criteria (Table 30). Compared to male teachers, female teachers demonstrated better knowledge of the English and mathematics curricula, used more teaching aids, and praised more than reprimanded. The number of female teachers who passed the English and mathematics content knowledge tests is twice that of male teachers, and the number of female teachers who meet the teacher competence logframe standard is three times that of male teachers who do so.

Part of the difference in teacher competence by gender is geographical: as discussed in the previous section, most female teachers are based in Lagos, and relatively few are in Kano or Jigawa. To explore this we examine the proportion of teachers meeting the overall competence standard by state and gender (the lower section of Table 30). Although the gap narrows, we still find within each state a higher proportion of female than male teachers reaching the overall competence standard.

Box 12: 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 reprimand during lesson observation; and
4) 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) English and mathematics content knowledge: scores at least 50% in both an English literacy and a mathematics test.

The CS2 overall report reports performance of teachers on these logframe standards in more detail and breaks down results by CS1 versus CS2, and across CS2 schools by Output 3 intervention groups. The CS2 state reports similarly delve into teacher competence in more detail at individual state level. The present report focuses on gender differences within CS2.

Table 30: Teacher competence in CS2 by gender groups

Teacher competence criteria (%)	Male	Female	Significant difference
Knowledge of English/maths curriculum	26.2	44.1	F
Use of one or more teaching aid	95	97.6	
Use of one or more teaching aid, excluding reading/writing/copying from blackboard	68.2	82.4	F
Praise more than reprimand	74.8	86.6	F
Assigns two or more individual/group tasks	46.2	59.5	F
English score (%)	39.7	53.4	F
Mathematics score (%)	55.5	67.7	F
Passes English and mathematics test	28	57.5	F
Competence score (CS1 version)	64	73.4	F
Meets teacher competence standard (CS1)	59.5	73	F
Competence score (CS2 version)	55	68	F
Teacher competence standard (CS2)	13.6	40.4	F
Teacher competence standard (CS2) by gender groups and states			
States	Male	Female	Significance
Enugu	39.8	59.0	F
Jigawa	6.9	14.1	
Kaduna	16.5	24.1	
Kano	9.6	25.0	F
Kwara	25.5	41.8	F
Lagos	42.3	59.5	

Note. F = significant difference in favour of females; M = significant difference in favour of males ($p < .05$)

Both male and female teachers appear to be more competent in ESSPIN than in non-ESSPIN schools, but there remain gender differences in both types of school (Table 31). However, for our measures of competent teaching based on lesson observations—use of teaching aids, praise and reprimand, and assigning individual and group tasks—there are large and significant differences in non-ESSPIN schools which largely disappear in ESSPIN schools. This suggests that these

aspects of teacher behaviour are things that most female teachers do anyway, but which are widespread among both male and female teachers within ESSPIN schools. Female teachers have higher scores than male teachers in English and mathematics tests in both ESSPIN and non-ESSPIN schools, but, again, the percentage-point gap is somewhat smaller in ESSPIN schools.

Table 31: Teacher competence in CS2, by gender and Output 3 intervention groups

	Control			Treatment		
Teacher competence criteria (%)	Male	Female		Male	Female	
Knowledge of English/maths curriculum	23.8	42.3	F	31.2	45.4	F
Use of one or more teaching aid	93.9	97.3		98.3	97.9	
Use of one or more teaching aid excluding reading/writing/copying from blackboard	64.1	77.4	F	79.8	86.8	
Praise more than reprimand	70	84.3	F	88.4	88.7	
Assigns two or more individual/group task	43.1	58.3	F	55.2	60.6	
English score (%)	37.2	51.5	F	46.8	55.3	F
Mathematics score (%)	52.4	64.1	F	64.8	71	F
Passes English and mathematics tests	23.6	53.3	F	41.3	61.5	F
Competence score (CS1 version)	61.7	72.5	F	70.7	74.2	
Meets teacher competence standard (CS1)	56	71.6	F	69.5	74	
Competence score (CS2 version)	52.3	65.8	F	63.5	70	F
Teacher competence standard (CS2)	10.4	36.4	F	23.7	44.3	

Note. F = significant difference in favour of females; M = significant difference in favour of males ($p < .05$)

8.3 Spatial and gender inclusion practices

Spatial and gender inclusiveness by teachers in classroom practices is discussed in more detail in section 4 above; here we disaggregate performance of the CS2 teacher cohort by gender on these two aspects (Table 32). Female teachers perform significantly better than their male counterparts on spatial inclusiveness, with 70% of female teachers engaging pupils from four or more areas of the classroom during lessons, compared to 50% of their male counterparts. Female teachers, on average, engage more than four areas of the classroom in lessons while male teachers engage slightly over three. On gender inclusiveness, female teachers come closer than male teachers to involving proportionate numbers of girls and boys.

Table 32: Spatial and gender inclusiveness by teachers in CS2 schools, by gender groups

Indicators (%)	Male	Female	Significant diff.
Involves boys/girls proportionately (within 10%)	47.4	53.5	
Gender equity score (0=completely unequal; 100=perfectly equal)	78.9	85.1	*
Involves pupils from four or more areas of the class	50.5	69.8	*
Number of zones participating in lessons	3.4	4.2	*

Note. * indicates a statistically significant positive/negative difference ($p < .05$); blank cells indicate no statistically significant difference.

9 Head teacher effectiveness: gender differences

In this section, we present head teacher effectiveness results by gender. Overall, 34% of CS2 schools were headed by female head teachers, with a wide variation between northern and southern states. Across all six states, female head teachers appear to be performing significantly better than their male counterparts on several criteria: overall, the number of female head teachers who meet the effectiveness standard is more than twice that of male teachers, and this result holds for ESSPIN schools too. The pattern of female head teachers outperforming male counterparts is, however, not mirrored uniformly at the state level.

9.1 Gender composition of head teachers in CS2

The gender composition (weighted) of head teachers in CS2 schools is presented in Table 33 below. Overall, 34% of the 728 schools visited in CS2 are headed by female head teachers. As expected, there is wide variation between northern and southern states, and even within these states. In the north, Jigawa had 1% schools headed by a female head teacher, with Kano and Kaduna at 6% and 30% respectively. In the south (and overall), Lagos has the highest proportion of schools headed by female head teachers (86%), followed by Enugu and Kwara.

Table 33: Percentage of schools headed by female head teachers in CS2 sample

	Female
Jigawa	1%
Kaduna	30%
Kano	6%
Enugu	56%
Kwara	45%
Lagos	86%

9.2 Head teacher effectiveness criteria

The ESSPIN logframe defines head teacher effectiveness in terms of seven criteria (Box 13). These reflect both activities by the head teacher and behaviour across the teachers and pupils, such as agreement on what time the school opens (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 13: Logframe standard for head teacher effectiveness

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

1) Carries out two or more lesson observations in the past two weeks;
2) Holds four or more professional development meetings since the start of the 2011–12 or 2013–14 school year (note that the survey took place more than nine months into the school year);
3) School has a teacher attendance book and the head teacher recalls at least two actions taken to promote teacher attendance;
4) Clear school opening time: more than 50% of pupils sampled agree on the school opening time and more than 50% of teachers sampled agree on the school opening time;
5) More than 50% of classes are in their classroom with their teacher within 30 minutes of the 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. between 30 and 40 minutes long).

The CS2 overall report examines head teacher effectiveness with regard to these criteria, over time and between ESSPIN and non-ESSPIN schools. The CS2 state reports delve into head teacher effectiveness in more detail at the state level. The present report focuses specifically on differences between male and female head teachers.

Across all CS2 schools and all six states together (Table 34), female head teachers appear to be performing significantly better than their male counterparts for four of the seven logframe head teacher effectiveness criteria. Compared to male head teachers, a higher percentage of female head teachers carried out two or more lesson observations in the past two weeks (criterion 1); held four or more professional development trainings (criterion 2); took appropriate action to improve teacher attendance (criterion 3); and were heading schools where the majority of the classes started within 30 minutes of school opening time (criterion 5). On criterion 7 (more than 50% of lessons observed finished within five minutes of a standard lesson of 35 minutes' duration), female head teachers are performing significantly worse than their male counterparts—the reason for their worse performance on this criterion is unclear and some measurement error here cannot be ruled out. The number of female head teachers who meet the effectiveness standard (5/7 criteria met) is more than twice that of male teachers, and this difference is significant.

The preceding sections showed that part of the gender difference in teacher competence actually reflected differences between the states where male and female teachers were most concentrated. A similar pattern applies for head teachers. Female head teachers are predominantly in Lagos, and very few are in Jigawa or Kano (see Table 33 above). However, when we break the results down by state (bottom part of Table 34), there remain large gender gaps in all states except Kano, although there much variation within each state and the gender difference reaches statistical significance only in Enugu. In Jigawa, we found only one female head teacher, and in Kano 10; in Lagos, by contrast, there were only 15 male head teachers. Our estimates of gender-specific statistics for these three states may, therefore, not be very reliable.

Table 34: Head teacher effectiveness in CS2 schools, by gender

Head teacher effectiveness criteria	Male	Female	Significant diff.
(1) lesson observations (%)	15.4	42	F
(2) professional development meetings (%)	14.1	50.4	F
(3) action on teacher attendance (%)	50.2	66.5	F
(4) clear opening time (%)	54.7	50.5	
(5) in class on time (%)	55.8	89.8	F
(6) appropriate morning break (%)	77.8	67.5	
(7) appropriate lesson length (%)	55.2	33.7	M
Number of criteria fulfilled (/7)	3.2	4	F
Effective head teacher (5/7 criteria met)	15.5	38.2	F
Head teacher effectiveness standard by gender and state			
State	Male	Female	Significance
Enugu	9.2	30.3	F
Jigawa	9.1	1 female HT	N/A
Kaduna	12.6	18.6	
Kano	19.4	13.7	
Kwara	17.7	29.8	
Lagos	58.9	67.7	

Note. F = significant difference in favour of females; M = significant difference in favour of males ($p < .05$). Estimates for female head teachers in Kano, and male head teachers in Lagos, are based on small sub-samples and so may not be reliable. As there was only one female head teacher in Jigawa (who met the effectiveness standard), we do not present any estimate for this sub-group.

The gender difference in head teachers' effectiveness is found in both ESSPIN and non-ESSPIN schools (Table 35), and by some measures the gender gap seems larger in ESSPIN schools. Nearly half of male head teachers in ESSPIN schools meet the effective head teacher standard, compared to only 22% of male head teachers. In non-ESSPIN schools, only 14% and 19% of male and female teachers, respectively, meet the same standard. This suggests that ESSPIN might be having larger effects on female head teachers, but as already noted, it may also represent geographical variation since disproportionate numbers of female head teachers are based in Lagos and very few in Jigawa or Kano.

Table 35: Head teacher effectiveness in CS2 schools, by intervention categories and gender

	Non-ESSPIN school			ESSPIN school		
Head teacher effectiveness criteria	Male	Female		Male	Female	
(1) lesson observations (%)	13	32	F	25	49	F
(2) professional development meetings (%)	11	27	F	29	64	F
(3) action on teacher attendance (%)	51	69		45	65	F
(4) clear opening time (%)	55	53		52	49	
(5) in class on time (%)	53	93	F	69	88	F
(6) appropriate morning break (%)	79	50	M	73	78	
(7) appropriate lesson length (%)	59	36		39	33	
Number of criteria fulfilled (/7)	3	3.6	F	3.3	4.2	F
Effective head teacher (5/7 criteria met)	14	19		22	49	F

Note. F = significant difference in favour of females; M = significant difference in favour of males ($p < .05$)

9.3 Actions to improve attendance

Actions taken by head teachers to improve pupil attendance have been discussed in detail in section 4 above; here we disaggregate these results for the CS2 head teacher cohort by gender (Table 36). Overall, female head teachers could list more actions they had taken to improve pupil attendance (2.6) than their male counterparts (2.2), and this difference is significant. They are twice as likely to say they have implemented solutions to improve pupil attendance suggested by teachers, pupils and parents, and four times as likely to say they have improved the quality of teaching and learning as a means of improving pupil attendance. However, with regards to involving SBMCs for finding reasons for pupil absenteeism, they perform worse than their male counterparts.

Table 36: Head teachers' actions to improve pupil attendance in CS2, by gender

% of head teachers who took the following actions	Male	Female	Significant diff.
Keep up-to-date registers	46.6	34.4	
Involve SBMC in finding reasons for non-attendance	43.4	27.1	M
Discuss with teachers, pupils or parents about reasons for non-attendance	73.2	70.2	
Implement suggested solutions for non-attendance	6.3	13.8	F
Address issues such as bullying and punishment	3	4.6	
Improve quality of teaching and learning	4.4	17	F
Average number of actions taken	2.2	2.6	F

Note. F = significant difference in favour of females; M = significant difference in favour of males ($p < .05$)

10 Conclusion

This report has examined a wide range of indicators—inclusive practices in schools, SBMC functionality and inclusiveness, and differences in education outcomes along various dimensions of marginalisation—in CS1 and CS2, in 2012 and 2014. It has asked whether things were getting better over time overall and across the six states, and whether ESSPIN schools are currently doing better than non-ESSPIN schools.

ESSPIN schools are performing better than the control schools in a number of these areas, namely school inclusiveness, spatial inclusion in the classroom, SBMC functionality, women's inclusiveness, and children's inclusiveness. We find evidence that the poorest children are benefiting disproportionately from ESSPIN, and that ESSPIN schools have smaller wealth disparities than non-ESSPIN ones.

However, the overall situation across ESSPIN states remains well below the programme's logframe standards and may be deteriorating in some cases, such as classroom spatial inclusion, and head teacher actions to improve pupil attendance, while there is lack of any progress in women's inclusiveness in the management of schools. Improvements were seen over time in SBMC functionality and children's inclusiveness but both these indicators remain low overall. There are wide disparities in pupil's learning outcomes in rural vis-à-vis urban schools and, though these are less pronounced in ESSPIN schools, it is still significant. Similarly, pupils from 'poorer' backgrounds are performing starkly worse than their better-off peers and again, though this disparity is dampened in ESSPIN schools, it is still significant.

Thus, overall, ESSPIN schools appear to be performing better than non-ESSPIN schools in terms of inclusive practices in schools, classrooms and SBMC, and along various dimensions of marginalisation. However, as mentioned above, the overall standard in the six ESSPIN states remains low. Given the nature of the record-keeping that needs to be produced by schools during the Composite Surveys, ESSPIN schools' better performance in areas such as SBMC functionality, women's and children's inclusiveness, as ESSPIN's own research shows, may be due to targeted record-keeping training. More pertinently, contextual factors, such as persistent conflict and gender inequality, in some of the ESSPIN states in northern Nigeria may also be part of the explanation for slow or non-existent improvement between CS1 and CS2 in the dimensions that this report explores.

In conclusion, inclusive practices in schools, classrooms and SBMCs show major weaknesses across states, and disparities in education outcomes along various dimensions of marginalisation remain vast and challenging, though ESSPIN schools and teachers appear to be performing better than non-ESSPIN schools in most of these areas. This report has thus discussed the state of inclusion practices both within Output 4 schools and across all schools in the six programme states. While low levels of inclusion reflect persistent sociocultural patterns that are beyond the control of ESSPIN, recent large-scale roll-out of interventions has meant that the programme now has direct links with a very large number of schools in the six states. Much of this roll out happened in 2013–14 and so is unlikely to have started having a major impact by the time of our survey, near the end of the 2013–14 school year. Over the next year, it is hoped that the scale-up will start to push up the overall state averages as opposed to just the averages within ESSPIN schools.

References

Cameron, S. (2015). ESSPIN Composite Survey 2: Overall Report. Oxford, UK: OPM.

DFID. (2013). Extension to the Education Sector Support Programme in Nigeria (Aug 2014–Jan 2017): business case.

EPDC. (2011). Efficiency: A Study of Promotion-, Repetition-, and Dropout Rates among Pupils in Four Age-groups in Sub-Saharan Africa. Retrieved from http://www.epdc.org/sites/default/files/documents/Efficiency_A_Study_of_Promotion_Repetition_and_Drop-out_Rates.pdf on 24 January 2015.

ESSPIN. (2013). ESSPIIN Overall findings and technical report. Report number: ESSPIN 060.

ESSPIN Community Engagement and Learner Participation team. (2014). Report of SMO/SUBEB Data 2012–14 for Annual Review. ESSPIN/UKAID.

Filmer, D., and Pritchett, L. H. (2001). Estimating wealth effects without expenditure data—or tears: an application to educational enrollments in states of India. Demography, 38(1), 115–132.

Lewin, K., and Ricardo, S. (2011). Changing patterns of access to education in Anglophone and Francophone countries in sub-Saharan Africa: is education for all pro-poor? (Research Monograph 52). CREATE: Centre for International Education, University of Sussex.

Lloyd, C. B. (2012). The demography of the classroom in the primary grades; Patterns of enrolment by age and implications for early learning. Hewlett Foundation.

Megill, D. (2014) Final sample design and weighting procedures for ESSPIN Second Composite Survey in Nigeria. Unpublished note written for ESSPIN and Oxford Policy Management.

Mont, D. (2014). Monitoring progress in education among individuals with disabilities. Retrieved from https://efareport.wordpress.com/2014/12/04/monitoring-progress-in-education-among-individuals-with-disabilities/

National Bureau of Statistics. (2009). Nigerian ASC.

National Bureau of Statistics. (2013). Nigerian ASC.

Ndaruhutse, S. (2008). Grade repetition in primary schools in Sub-Saharan Africa: an evidence base for change. Retrieved from http://birbhum.gov.in/DPSC/reference/83.pdf

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

Pritchett, L., and Beatty, A. (2014). Slow down, you're going too fast: Matching curricula to student skill levels. International Journal of Educational Development.

Taylor, N., Mabogane, T., Shindler, J., Akoobhai, B. (2010). Seeds of their struggle: the features of under- and overage enrolment among grade 4 learners in South Africa. CREATE: Centre for International Education, University of Sussex. Retrieved from http://r4d.dfid.gov.uk/Output/184883/

Annex A Inclusion practices in schools and SBMCs: Results by ESSPIN's Output 3 interventions

A.1 School inclusiveness

Table 37 shows that schools in which we would expect an improvement due to ESSPIN Output 3 intervention are significantly more inclusive than non-ESSPIN schools in terms of activities to improve access for disadvantaged children (criterion 2), use of different assessment methods (criterion 3), and in terms of the overall inclusiveness score and the proportion of schools meeting standards. Overall, 25% of ESSPIN schools and only 8% of non-ESSPIN schools fully meet the inclusiveness standard. Teachers in ESSPIN schools used more assessment methods, involved children from more parts of the classroom in lessons, and came closer to ensuring equitable participation of boys and girls in lessons.

Table 37: School inclusiveness in CS2 by Output 3 intervention group

School inclusiveness criteria	(1) no expected improvement	(2) expected improvement	
Inclusiveness criteria			
(1) three or more actions to improve attendance	39.4	38.3	
(2) two or more activities in SDP to improve access for disadvantaged children	4.4	32.1	+
(3) >50% of teachers use two or more assessment methods	55.0	81.4	+
(4) >50% of teachers spatially inclusive and >50% are gender inclusive	20.6	30.4	
Overall inclusiveness standard			
Number of inclusiveness criteria fulfilled (/4)	1.2	1.8	+
Inclusiveness score	61.3	69.7	+
School partially meets inclusiveness standard (2–4 criteria out of 4)	40.1	63.6	+
School fully meets inclusiveness standard (3–4 criteria out of 4)	**8.0**	**25.1**	**+**
Detailed			
Number of actions to improve attendance	2.3	2.3	
Number of activities on access for disadvantaged children	0.1	1.1	+
Average number of assessment methods used	1.1	1.9	+
Average number of zones participating in lessons	3.5	3.9	+
Average gender equity score (0=completely unequal; 100=perfectly equal)	80.6	86.7	+

Note: + indicates a significant positive difference between non-ESSPIN and ESSPIN schools. The gender equity score for a teacher is $\mathbf{100} - \mathbf{100} \times \boldsymbol{abs}(\frac{g}{g+b} - \frac{G}{G+B})$ 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.

A.2 Head teachers' actions to improve pupil attendance

Focusing on CS2 schools (Table 38), actions taken by head teachers to improve pupil attendance were not significantly different in schools where we would expect an improvement due to ESSPIN Output 3 intervention, in relation to the comparison group. Fewer head teachers in the schools where we would expect an improvement reported keeping up-to-date registers as a means of curbing pupil absenteeism, while more of them reported improving quality of teaching and learning. Overall, for a majority of the actions listed, the percentage of head teachers using these as means to improve pupil attendance was not significantly different across the two categories of school.

Table 38: Head teachers' actions to improve pupil attendance in CS2 schools by Output 3 intervention groups

% of head teachers who took the following actions	No expected improvement	Expected improvement	
Keep up-to-date registers	49	31.5	-
Involve SBMC in finding reasons for non-attendance	40.9	39	
Discuss with teachers, pupils or parents about reasons for non-attendance	71.5	75.6	
Implement suggested solutions for non-attendance	7.8	7.1	
Address issues such as bullying and punishment	2.8	4.7	
Improve quality of teaching and learning	4.6	12.6	+
Average number of actions taken	2.3	2.33	

A.3 Spatial and gender inclusiveness

Do teachers in ESSPIN schools demonstrate more gender and spatial inclusiveness than their counterparts in non-ESSPIN schools? We disaggregate the CS2 teacher sample by the categories discussed above, i.e. non-ESSPIN school, ESSPIN school, and ESSPIN-trained teacher (Table 39). Teachers who are in ESSPIN schools but have not been directly trained by ESSPIN (second column) perform significantly better than teachers in non-ESSPIN schools on the spatial inclusiveness criteria. They perform better than non-ESSPIN counterparts on gender inclusion too, but this difference is not significant. ESSPIN-trained teachers also perform better than non-ESSPIN school teachers, but this is significant only for the gender equity score (under gender inclusiveness) and for the number of zones participating during a lesson (under spatial inclusiveness).

Table 39: Spatial and gender inclusiveness by teachers in CS2 schools by Output 3 intervention group

Indicators (%)	Non-output 3 school	Output 3 school		Output 3 trained	
Involves boys/girls proportionately (within 10%)	50.7	51.5		57.6	
Gender equity score (0=completely unequal; 100=perfectly equal)	80.8	83.2		87.5	+
Involves pupils from four or more areas of the class	55.5	64.9	+	66.3	
Number of zones participating in lessons (#)	3.6	3.9	+	4.2	+

A.4 SBMC functionality

In terms of differences between ESSPIN and non-ESSPIN schools in CS2, there were significant differences in the expected direction across all nine criteria (Table 40). Overall, only 13% of non-ESSPIN schools, but 62% of ESSPIN schools, met the standard for a functional SBMC.

Table 40: SBMC functionality in CS2 by Output 3 intervention groups

SBMC functionality criteria	(1) no expected improvement	(2) expected improvement	
(1) two or more meetings this school year	13.7	62.6	+
(2) conducted awareness raising	39.2	71.1	+
(3) addressed exclusion	32.5	59.8	+
(4) networked	47.5	75.8	+
(5) interacted with LGEA	16.9	34.0	+
(6) has a women's committee	14.8	57.7	+
(7) has a children's committee	10.4	51.8	+
(8) contributed resources for school	47.8	73.1	+
(9) chair visited school three or more times	5.6	39.0	+
Standard G: functioning SBMC	17.1	67.0	+
Number of SBMC functionality criteria met	2.5	5.3	+
Additional criteria			
Action for commonly excluded groups	23.2	27.1	
Raised issue of children's exclusion	15.3	29.8	+

A.5 SBMC women's inclusiveness

Focussing on CS2 schools, there were large differences in women's inclusiveness between ESSPIN and non-ESSPIN schools; these are significant and in the expected direction across the four criteria (Table 41). Nearly half of ESSPIN schools, but very few non-ESSPIN schools, met the standard for women's inclusiveness.

Table 41: SBMC women's inclusiveness in CS2 by Output 3 intervention group

SBMC women's inclusiveness criteria	(1) no expected improvement	(2) expected improvement	
(1) at least one woman attended two or more meetings (%)	4.6	49.5	+
(2) a female member raised an issue (%)	16.8	66.6	+
(3) an issue raised by female member led to action (%)	4.2	39.7	+
(4) a women's committee meets (%)	13.3	58.3	+
Number of criteria met	0.3	2.1	+
Meets standard (3/4 criteria)	2.4	48.3	+

A.6 SBMC children's inclusiveness

As with women's inclusiveness, there are large positive differences between ESSPIN and non-ESSPIN schools (Table 42). Overall, 18% of ESSPIN schools, but fewer than 2% of non-ESSPIN schools, met the standard for SBMC children's inclusiveness.

Table 42: SBMC children's inclusiveness in CS2 between Output 3 intervention groups

SBMC children's inclusiveness criteria	(i) no expected improvement	(ii) expected improvement	
(1) a child attended two or more meetings (%)	2.3	25.2	+
(2) a child raised an issue (%)	11	44.5	+
(3) an issue raised by child led to action (%)	2.5	19.1	+
(4) a children's committee meets and it has a trained facilitator (%)	7	32.6	+
Number of criteria met	0.2	1.2	+
Meets standard (3/4 criteria) (%)	1.6	17.7	+

Annex B ESSPIN Output 3 intervention categories

The table below shows the ESSPIN Output 3 interventions delivered to date in each state. In order to make the variation in interventions across and within states manageable for analysis, each combination of interventions was categorised as *none*, *minimum*, *medium*, or *maximum*, according to the number of years of continuous intervention.

	category	2009–10			2010–11			2011–12				2012–13			2013–14			
		L	T	SV	L	T	SV	L	T	SV		L	T	SV	L	T	SV	
Enugu	none																	
	minimum														6	3	9	
	medium (1)							6	3	9		6	3	9	3		9	
	medium (2)											6	3	9	6	3	9	
Jigawa	none																	
	minimum														6	3	9	
	medium (1)											6	3	9	6	3	9	
	medium (2)	5*	5*	9*	10*	5*	9*											
Kaduna	none										**CS1**							**CS2**
	minimum														6	3	9	
	medium (1)											6	3	9	6	3	9	
	medium (2)							6	3	9		6	3	9	3		9	
	maximum	5*	5*	9*	10*	5*	9*	6	3	9		6	3	9	3		9	
Kano	minimum														9	9	9	
	medium	5*	5*	9*	10*	5*	9*								9	9	9	
Kwara	medium	6	3	30	6	3	30			30				30	6	3	30	
Lagos	medium (1)											6	3	9	6	3	9	
	medium (2)							6	3	9		6	3	9	3		9	
	maximum	5*	5*	9*	10*	5*	9*	6	3	9		6	3	9	3		9	

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

Annex C ESSPIN Output 4 intervention categories

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

	Category	2010–11			2011–12				2012–13			2013–14				De-facto phases for analysis
		S	P	M	S	P	M		S	P	M	S	P	M		
Enugu	none															control
	2c											7		4		post-CS1
	2b								7		4	r	6	4		post-CS1
	1 / 2a				7		4		r		4		6			pre-CS1
Jigawa	none															control
	2b											2				control
	2a											7		4		post-CS1
	1	7		4	r		4			6	4*			4*		pre-CS1
Kaduna	none							CS1							CS2	control
	2a				7		4		r		4		6	4*		pre-CS1
	1	7		4	r		4			6	4*			4*		pre-CS1
Kano	none															control
	2b											r	6			
	2a											7		4		post-CS1
	1				7		4		r	6	4			4*		pre-CS1
Kwara	none															control
	2b											7				post-CS1
	2a											4		2		post-CS1
	1				7		4		r		4		6	4*		pre-CS1
Lagos	2b											7		4		post-CS1
	1 / 2a				7		4		r		4	7		4		pre-CS1

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

Annex D Overlaps between ESSPIN outputs 3 and 4

Further complexity in the classification of schools is introduced by the fact that the intensity of outputs 3 and 4 is not always the same. As Table 43 below shows, the agreement between intensity and timing of roll-out of outputs is not perfect, which necessitates separate disaggregation of results by individual outputs in the analysis and discussion.

Table 43: Overlap between outputs 3 and 4

	Output 4	Output 3
Enugu	control	control
	post-CS1	post-CS1
	pre-CS1	pre-CS1
Jigawa	control (1)	control
	control (2)	post-CS1
	post-CS1	post-CS1
	pre-CS1	pre-CS1
Kaduna	control	control
	pre-CS1	post-CS1
	pre-CS1	pre-CS1
Kano	control	post-CS1
	post-CS1	post-CS1
	pre-CS1	post-CS1
Kwara	control	pre-CS1
	post-CS1	pre-CS1
	pre-CS1	pre-CS1
Lagos	post-CS1	post-CS1
	pre-CS1	pre-CS1

In order to manage the number of disaggregations presented in this report, we use only the Output 4 classification to report results on inclusive practices in schools and SBMCs, which arise more directly from Output 4 activities, in Part A; corresponding classification of indicators by Output 3 are tabulated in Annex A. Part B discusses more broad-based differences in education outcomes at pupil, teacher and school levels, then reports results by ESSPIN's Output 3 categories discussed above.