ESSPIN Composite Survey 3

Gender and inclusion report

Rachita Daga

Acknowledgements

Oxford Policy Management Limited	Level 3, Clarendon House	Tel	+44 (0) 1865 207 300
	52 Cornmarket St	Fax	+44 (0) 1865 207 301
	Oxford OX1 3HJ	Email	admin@opml.co.uk
Registered in England: 3122495	United Kingdom	Website	www.opml.co.uk

Executive summary

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

This report presents findings from the first, second and third rounds of the Composite Survey (CS1, CS2 and CS3), conducted in 2012, 2014 and 2016, respectively. 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 understand how inclusive practices in schools and SBMCs in Education Sector Support in Nigeria (ESSPIN) states are changing over time and 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 a significant and positive difference between schools that receive ESSPIN interventions, and schools that do not receive ESSPIN interventions. Schools that receive ESSPIN interventions perform significantly better than non-ESSPIN schools on almost all school inclusiveness criteria, as well as on the number of inclusiveness criteria fulfilled, and the schools that partially/fully met the inclusiveness standard. ESSPIN schools are also more inclusive in terms of activities to improve access for disadvantaged children, and the use of different assessment methods, as compared to non-ESSPIN schools. However, the inclusiveness in schools between CS2 and CS3 has no clear trend, with inclusiveness measured by certain indicators improving during this period, whereas others worsen. There is no significant change in the schools that fully meet the inclusiveness criteria, but there is a decline in the number of schools that partially meet the inclusiveness criteria in CS3 as compared to CS2, indicating that a greater number of schools have failed to meet the inclusiveness criterion in CS3 as compared to CS2. Between CS1 and CS3, however, we find that there is a significant and positive improvement in school inclusiveness, both in terms of the overall inclusiveness score, as well as the number of schools that fully and partially meet the inclusiveness standard.

Head teachers' actions to improve pupil attendance: A greater percent of the head teachers surveyed in CS3 involved teachers in finding reasons for non-attendance and implementing the suggested solutions for non-attendance, as compared to head teachers surveyed in CS2. However, there is no significant difference between the number of actions taken by head teachers to address pupil attendance between CS2 and CS3. Within CS3 schools, there was also no significant difference between head teachers in ESSPIN schools and those in non-ESSPIN schools for most indicators, including the average number of actions taken.

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. Teachers in CS3 were, on average, less spatially inclusive than those in CS2; but in CS3 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 between CS2 and CS3

SBMC functionality and actions to support inclusion: SBMC functionality improved significantly between CS2 and CS3 for almost all the criteria, as well as the number of schools that met the logframe standard for SBMC functionality. Between CS1 and CS3, we find that there is a significant and positive improvement in eight of the nine SBMC functionality criteria, the number of schools meeting the SBMC functionality standard, the number of standards met by schools, as well as SBMCs raising issues of children's exclusion. Altogether around 5100 more schools have functioning SBMCs in 2016 than in 2012. The average school in CS3 met a little over 4.5 criteria for SBMC functionality, whereas those in CS2 met 3 criteria – a significant difference. Within CS3, ESSPIN schools performed significantly better than non-ESSPIN schools, for almost all criteria, as well as the number of criteria met, and the number of schools meeting the SBMC functionality logframe standard.

Participation of women in SBMCs: SBMCs in CS3 were significantly more inclusive of women as compared to SBMCs in CS2. SBMCs in CS3 were also significantly more inclusive than SBMCs in CS1. However, there is no significant difference between the number of SBMCs meeting the participation of women in SBMC standard between CS1 and CS3. An average school in CS3 met almost two criteria for women's inclusiveness in SBMCs, compared to schools in CS2, which met less than one criteria. This is in line with the scale-up in women and children's participation training after 2014. Schools that received interventions under ESSPIN performed significantly better than the control schools on all women's inclusiveness criteria.

Participation of children in SBMCs: There was a small but significant increase between CS2 and CS3 in the overall number of criteria met, as well as in the proportion of schools meeting the standard, which increased from 6% to almost 10%. There was also a significant and positive difference between the participation of children in SBMCs between CS1 and CS3 on all four criteria, as well as for the number of criteria met, and the percentage of SBMCs meeting the standard. However, there are large differences between ESSPIN and non-ESSPIN schools, with ESSPIN schools performing significantly better than non-ESSPIN schools. Between 19% and 25% of ESSPIN schools met the children's inclusiveness logframe standard, as compared to 4.4% of non-ESSPIN schools.

Overall, we also find that amongst ESSPIN schools the schools that received intervention after CS1, i.e. post-CS1 schools, have relatively better and higher outcomes as compared to those schools that received intervention prior to CS1, i.e. the pre-CS1 schools.

Part B: Differences in education outcomes by gender and background

Gender differences: We find that boys perform significantly better than girls on all tests, with the exception of the numeracy test for Grade 2 students. The same trend is observed for schools that received minimum intervention under ESSPIN Output 3, whereas the differences are less marked for schools that received medium and maximum intervention. There are variations between the states, with boys performing significantly better than girls in Kano, and girls performing significantly better than boys in Lagos. There is no clear trend for the other states.

Wealth: We find that an increase in the wealth index has a significant and positive impact on the performance of pupils. An increase in the amount of intervention received under ESSPIN Output 3 mitigates this effect to some extent, with the wealth gap in schools receiving maximum Output 3 intervention being relatively lesser than the corresponding gap in schools receiving minimum intervention. However, this is significant only for literacy tests and not for numeracy tests. Only 12% and 4% of the poorest pupils in Grade 4 meet the Grade 4 literacy and numeracy standards, respectively. 33.2% and 13.1% of the poorest pupils who meet the Grade 4 literacy and numeracy standards respectively come from schools that received medium or maximum intervention under

ESSPIN Output 3. Only 1.2% and 0.8% of the poorest pupils who meet the Grade 4 literacy and numeracy standards, respectively, come from schools that received minimum intervention.

Speaking a minority language: We find that pupils who predominantly speak a majority language (i.e. *not* Igbo in Enugu; Hausa in Jigawa, Kano, Kaduna; and Yoruba in Kwara) attain significantly higher results in the numeracy tests, as compared to students speaking minority languages. However, there is no significant difference for literacy tests between those who speak the majority language of the state and those who do not.

Disability: The CS3 pupil tests included a number of questions assessing various forms of physical disability among sampled pupils; children who were impaired were not asked to attempt parts of the test that would have been impossible for them to do without adaptation. In total, 51 pupils from the CS3 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.

Location: The test scores of the pupils are disaggregated by rural/urban location, as well as distance from Local Government Authority (LGA) headquarters. When we disaggregate the results by rural or urban location, we find that pupils from urban schools perform significantly better than pupils from rural schools, and this difference in performance of pupils increases from Grade 2 to Grade 4. When we disaggregate the performance of pupils in rural and urban schools on the basis of the amount of Output 3 intervention received we find that urban schools still perform significantly better than rural schools. When we disaggregate the results by distance from the LGA headquarters, we find that an increase in the distance from the LGA headquarters has a significant and negative impact on the performance of girls in Grade 4.

Age correctness for grade: 46.2% and 46.5% of the pupils in Grade 2 and Grade 4, respectively, are over-age, with almost 50% of all the pupils in Kaduna and Kano being over-age. A roughly equal percent of boys and girls are of equal ages, with the percent of over-age boys slightly exceeding that of over-age girls in Grade 4. A majority of the over-age pupils exceed the appropriate age by one year (Grade 2 pupils) or two years (Grade 4 pupils). Though the literature on over-age pupils remains divided, we find that over-age pupils perform significantly better than age appropriate pupils in all tests barring the Grade 2 literacy test. Over-age pupils perform significantly better than age appropriate pupils on all tests for schools receiving minimum intervention under ESSPIN Output 3. However, the difference in the performance decreases as the school receives more intervention, i.e. medium or maximum, under ESSPIN Output 3.

Teacher competence: Overall, 45% of CS3 teachers were female, with wide variation between northern (Jigawa: 12.7%) and southern states (Lagos, Enugu 80%+). 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 only mirrored in Enugu and Jigawa. The proportion of female teachers who passed both the teacher tests in English and mathematics is more than twice that of male teachers. Female teachers perform significantly better than their male counterparts on almost all logframe indicators and criteria for teacher inclusiveness for schools that received minimum and medium intervention, but there are no significant differences in the performance of male and female teachers in schools that received maximum intervention under ESSPIN Output 3. Female teachers are also more inclusive, both in terms of spatial inclusiveness as well as in regard to gender inclusiveness.

Head teacher effectiveness: In CS3 across all six states, female head teachers appear to be performing significantly better than their male counterparts: the number of female head teachers who met the effectiveness standard is almost three times that of male teachers, and this difference is significant. Female head teachers perform significantly better than male head teachers in

schools that receive minimum interventions, but the differences are less pronounced and less significant for schools that receive medium and maximum intervention. However, it should be noted that there are wide variations between the states, with states in the north having far fewer female head teachers than those in the south. Jigawa, for instance, has only one female head teacher.

Table of contents

List of tables, figures and boxes

List of abbreviations

ASC	Annual School Census
CAPI	Computer-assisted personal interviewing
CBO	Community-based organisation
CS1	Composite Survey 1 (2012)
CS2	Composite Survey 2 (2014)
CSO	Civil society organisation
EPDC	Education Policy and Data Centre
ESSPIN	Education Sector Support Programme in Nigeria
IRT	Item response theory
L2	Grade 2 literacy test
L4	Grade 4 literacy test
LGA	Local Government Area
LGEA	Local Government Education Authority
N2	Grade 2 numeracy test
N4	Grade 4 numeracy test
OPM	Oxford Policy Management
SBMC	School-Based Management Committee
SIP	School Improvement Programme
SMO	Social Mobilisation Officer
SSO	School Support Officer
SUBEB	State Universal Basic Education Board

1   Introduction

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

The Composite Survey aims to assess the effects of the integrated School Improvement Programme (SIP) and report on the quality of education in the six ESSPIN-supported states. The main purpose of the third Composite Survey (CS3) is both to provide insights into the changes over time in the six states where ESSPIN operates, and to evaluate whether ESSPIN is having an effect in the specific schools where its school improvement and community inclusion interventions have been applied. We are interested in a wide range of output indicators: teacher competence, head teacher effectiveness, school development planning, school inclusiveness, the functionality the participation of women and children in SBMCs, and the actions that SBMCs are taking to support inclusive education. Some of these same indicators are also combined to give an overall indicator of school quality. Finally, ESSPIN's impact is measured in terms of improved pupil learning outcomes, which we ascertain through test scores in numeracy and English literacy at Grades 2 and 4. CS3, which was conducted in 2016, aims to provide post-intervention data which can be compared to data from the first and second rounds of the survey (CS1 and CS2), collected in 2012 and 2014 respectively, in order to evaluate the extent of improvements in key indicators and gauge programme success.

The overall CS3 report presents findings from the CS3 and comparisons between CS1, CS2, and CS3, 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 is mostly descriptive and focuses on a sub-set of indicators.

The results presented in this report are based on cross-sectional analysis, and hence these results should be taken as indicative of change, but without attributing change. Hence, the results do not conclusively refer to programme effectiveness.[1] This report attempts to understand the following:

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

---

[1] The overall report (Cameron et al, 2016) can be referred to for further analyses based on difference-in-difference and matching methods.

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 capability of state and local governments for governance and management of schools; (iii) strengthened capability of primary schools to provide improved learning outcomes; (iv) and improved community engagement in school improvement.

As Figure 1 shows, under the fourth of these outputs (circled in red below), ESSPIN's community engagement and learner participation activities (hereafter, Output 4) aim to improve inclusive practices in basic education to enhance access to and quality of basic education, and therefore better 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:

- O4.1: SBMCs are functional 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;
- O4.2: Community and government organisations are better able to press for school improvement; and
- O4.3: 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 a minimum of 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.

Figure 1:   ESSPIN's community engagement and learner participation intervention
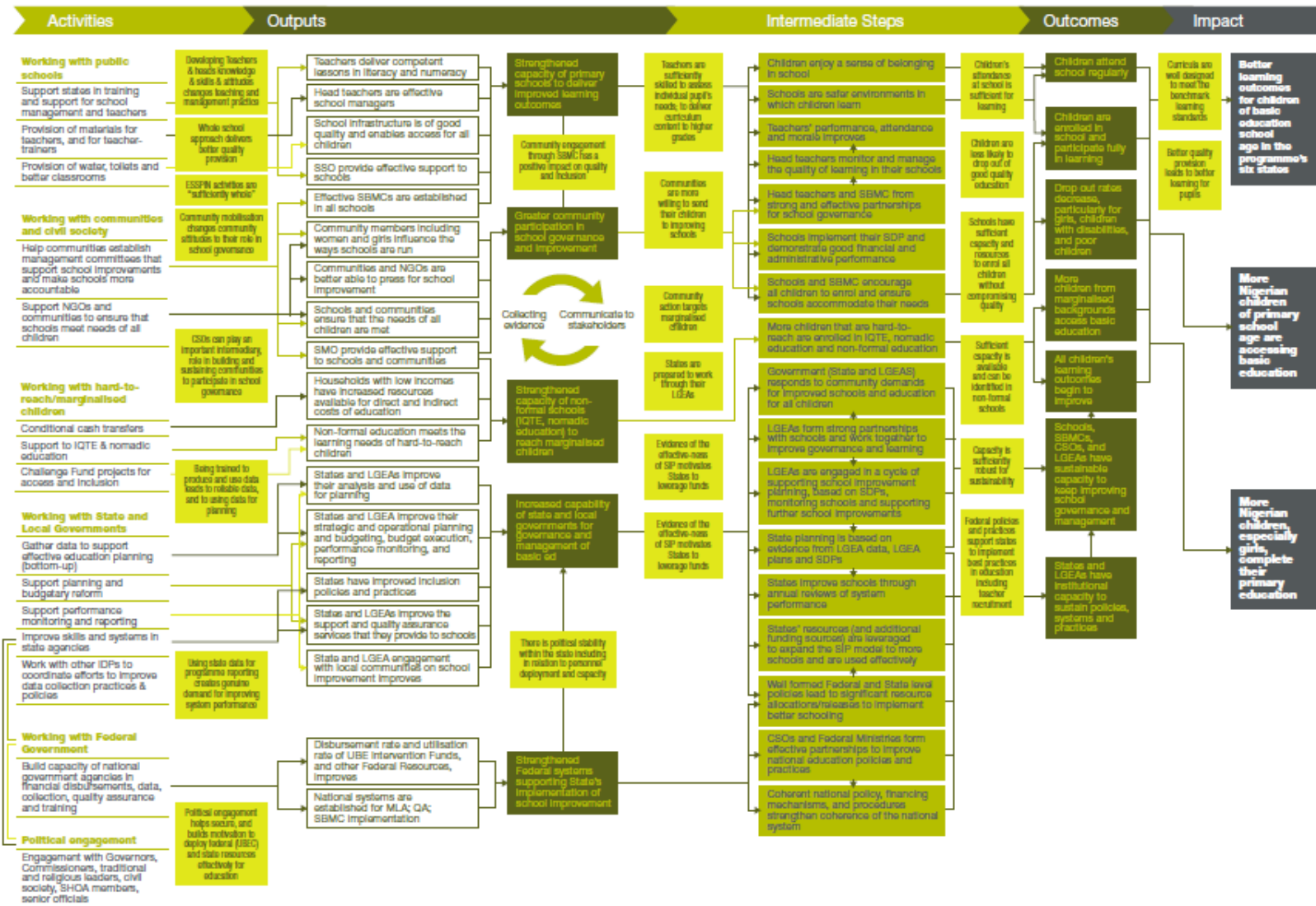
1.2   Coverage and scale-up since 2010/11

ESSPIN's activities under Output 4 have seen significant scaling up, particularly during the 2013/14 school year (Figure 2). There have been very few changes after 2013/14, with the exception of Jigawa, where Output 4 interventions were expanded to 918 schools in the form of SBMC training. While at 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]; while in Kaduna, 15% schools were covered initially and this has remained largely constant since 2012/13. We only consider schools that have received five or more days of Output 4 interventions as ESSPIN schools. Though the intervention was rolled out in 4,505 schools in Kano and 791 schools in Enugu after the completion of CS2, these schools have only received two days and one day of Output 4 intervention, respectively, and hence these are treated as control schools for the purpose of this analysis. After CS2, there has been a move towards deepening coverage, with the schools already receiving ESSPIN Output 4 interventions being exposed to mentoring visits and women and children training.

Figure 2:   Number of schools receiving ESSPIN Output 4 interventions
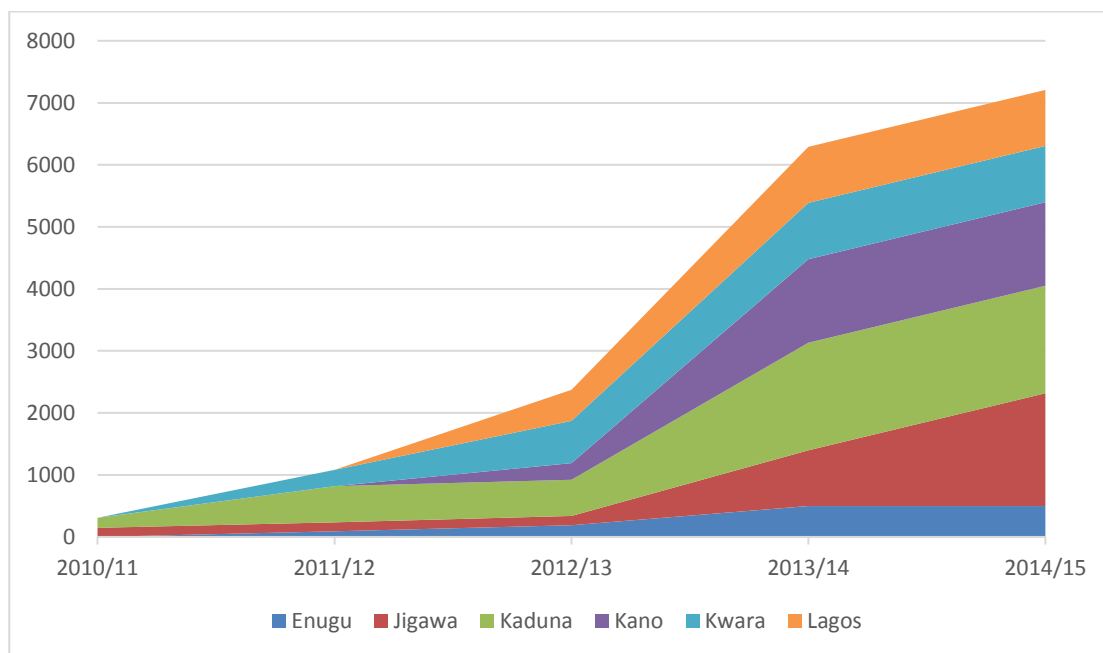
Source: Author's calculations based on intervention information provided by ESSPIN. 2015/16 is not added to the above graph because there were no new schools that received more than five days of intervention post-2014/15.

1.3   Structure of this report

The rest of this report is organised as follows. Section 2 describes the sample and its coverage, CS3 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, actions of head teachers to improve pupil attendance, and spatial and

---

[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, having completely restructured its SBMC development model over the last year from a cluster-based system to the ESSPIN-supported school-based system.

gender inclusive behaviours of teachers. In Part A, we also discuss the extent to which SBMCs are functional and inclusive of women and children (Section 5). In Part B, we discuss differences in education outcomes pertaining to pupils, teachers and head teachers by gender and background (Section 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 0), we discuss differences in terms of gender and particularly teachers' spatial and gender inclusive behaviour in the classroom. In Section 8.4, gender differences in head teacher effectiveness are discussed. Section 9 concludes.

2 Methods

2.1 Sample and weights

2.1.1 Sample design

In CS3, all the schools visited in CS2 were visited again, with the intention of collecting data that would 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. Due to the high variability in the types of schools in the different states, it is difficult to construct a representative sample. Hence, the sample design prioritised the ability to draw conclusions across the six states, given that it will not be possible to obtain statistically significant estimates within each state. The sampling design also incorporated the key aims of the study – to analyse change over time and differences between ESSPIN and non-ESSPIN schools.

Within each school, the survey team conducted interviews with the head teacher, the SBMC chairperson or deputy, teachers and pupils. Teachers that taught religious subjects were excluded from the survey. In order to assess changes over time and the rate of change in teacher competence and test results with more precision, the survey teams attempted to find the six teachers interviewed during CS2, using their photographs and name information, and to interview them. However, it was not always possible to find the teachers sampled in CS2, and, as a result of this, the sample had to be topped up with teachers randomly selected from the head teacher's register. Fieldwork teams asked the head teachers to complete the registers in cases where such a register was not available on the day of the visit. Team supervisors entered the number of eligible teachers into the computer-assisted personal interviewing (CAPI) system, which then randomly selected the additional top-up teachers to be sampled.

Pupil sampling in CS3 was slightly different from CS2. In CS2, pupils were randomly selected from the attendance register or through manual counting in class. However, in CS3 the same method could not be followed since it did not allow children to be linked to teachers. In order to address this limitation, teachers were sampled first in CS3. Then, pupils were randomly selected from the arms taught by the sampled teachers.

As in CS2, 16 pupils[4] were sampled per school in CS3, with a sample of four pupils per school for each of the tests (Grade 2 literacy, Grade 2 numeracy, Grade 4 literacy and Grade 4 numeracy). While sampling them, we gathered information from the head teacher on which teachers teach which arms of Grade 2 and 4, filling in an 'arm eligibility grid'. If there were no children in (for example) Grade 2 taught by the sampled teachers, we selected the pupils from all the arms of Grade 2. If there were fewer than 10 children in the selected arms of Grade 2, then we sampled from the selected arms first and then topped up randomly from the remaining arms. Sampling within the arms selected was done with cards, which were either blank or contained a number. This ensured that the sampling was done from the learners actually present in the class rather than from the potentially incomplete pupil register.

Although it would have been useful to trace the same pupils over time, this was not seen as feasible. This is because, for the children sampled in CS1, we did not have sufficient information to

---

[4] This was true for all schools except those in Lagos, where our sample of Grade 4 pupils needed to be large enough to allow for a comparison with private schools (which were surveyed as part of a separate exercise at the same time). We therefore increased the sample to eight literacy and eight numeracy tests for Grade 4 in Lagos. Therefore, a total of 24 pupils were sampled in each school in Lagos, eight pupils in Grade 2 and 16 pupils in Grade 4.

trace them back. Therefore, a random sample within each school in CS3 as well as in previous rounds of the survey was collected.

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

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 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, CS2 and CS3 schools, teachers, and pupils. A smoothing technique was also applied to reduce the variability of the weights and to avoid the design effects problem encountered in the CS1 analysis (see Megill, 2014). Most of the following analysis applies weights to sample statistics calculated within each round and intervention group. This 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 CS3 consisted of 735 schools, which were successfully visited by the field team. Head teachers and SBMC chairs/deputies were interviewed in all the schools visited. The number of teachers interviewed fell short of the targeted sample size in all states. The primary reason for this was that many schools had fewer than six teachers on their staff roll who teach non-religious subjects to Primary Grade 2 (P2) or Primary Grade 4(P4). In all six states, sample coverage of eligible teachers was above 99.5%. Similarly, the number of pupils assessed fell slightly short of the targeted number because some schools had fewer than eight pupils in P2 or P4 (fewer than 16 pupils in the case of P4 in Lagos). There were some further minor differences between the numbers of teachers interviewed and those who were tested or whose lessons were observed because some teachers did not give their consent to be tested or observed; and because some teachers reportedly became ill and so could not complete the test or lesson.

Table 1: **Sample coverage in CS3**

	Schools			Teachers			Pupil tests		
	Intended sample	Actual	Interview	Less. obs	Tests	L2	L4	N2	N4
Enugu	105	105	547	537	546	391	397	377	391
Jigawa	105	105	462	462	461	408	411	406	410
Kaduna	140	140	653	653	654	552	542	549	539
Kano	175	175	838	837	825	696	671	691	670
Kwara	105	105	495	492	498	385	380	373	366
Lagos	105	105	592	591	592	415	807	413	805
Total	735	735	3587	3572	3576	2847	3208	2809	3181

2.2   Training, pilots and fieldwork model

Fieldwork for CS3, including the pupil tests, was conducted using CAPI during May–July 2016. Children were given a printed pupil book to read and write in. The interviewers held a tablet computer, prompting them with the questions the children were to be asked orally, giving instructions on administration of the different test items. This included 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 pre-tested over two days in Abuja during April 2016, after training state coordinators and monitoring officers. The state coordinators and monitoring officers collected the data using CAPI after they had been trained on the instruments. Instruments were revised through consultations with ESSPIN and state coordinators.

Table 2 lists the instruments used in CS3, together with the indicators relevant to outcomes, outputs or impact that were gathered from each instrument. The instruments were also used to gather intervention information, such as whether individual teachers had received ESSPIN training or not, and pupil-level information on socioeconomic status, age, language spoken at home, and gender. The data gathered in general allow more detailed analysis than that presented in this report, some of which is presented in the six state-level reports and the overall 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 CS3 does leave some possibility of measurement error in comparisons between CS1, CS2 and CS3. The priority for CS3 was to ensure consistent and manageable data collection within CS3 by setting clearer guidance for data collectors through detailed data collection manuals, applying 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 and CS2, 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, since changes in measurement are consistent across ESSPIN and non-ESSPIN schools, they should not affect any within-CS3 comparisons.

Table 2:   Instruments used in CS3

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

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, SBMC chair persons, and through lesson observations. 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 comparing the change in indicators from between CS2 (2013/14) and CS3 (2016/17). ESSPIN Output 4 interventions have been expanded to more schools since CS2, but there are still schools that have yet to receive any interventions under Output 4, or that have only just started receiving Output 4 interventions with the magnitude being less than five days. In addition to this, there are also large variations in the schools that have received the Output 4 interventions relating to mentoring visits and women and children.

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

   o *pre-CS1 schools 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 CSO 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; additional support to enhance the participation of women and children in school-based management and to support SBMC forums at LGEA level.

   o *post-CS1 schools:* schools added as part of the roll-out of Output 4 interventions during 2012/13, 2013/14, 2014/15 and 2015/16. However, there were hardly any new roll-outs after 2013/14, with, instead, a greater focus on increasing the amount of interventions received by the schools already receiving Output 4 interventions.

   o *control schools:* schools which have not yet received any Output 4 intervention or have received less than five days of intervention under Output 4.

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). We test significance at the 5% confidence level. 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.

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 that schools with more ESSPIN intervention were significantly more inclusive than other schools. This is also true for spatial and gender inclusiveness, with ESSPIN schools performing significantly better than non-ESSPIN schools. Amongst ESSPIN schools, we find that schools that started receiving interventions post-CS1 have more positive and significant results as compared to schools receiving interventions prior to CS1. Analysing inclusiveness over time, we find that there is no significant change in overall inclusiveness between CS2 and CS3. The schools surveyed in CS3 perform worse than the schools surveyed in CS2 on the number of actions taken on attendance and the percentage of schools that partially meets the inclusiveness criteria, and perform better than CS2 schools on actions in the SDP for disadvantaged children. The average number of actions taken by head teachers to improve pupil attendance is lower in CS3 compared to CS2. On average, there is no significant difference between CS3 and CS2 schools on spatial and gender inclusiveness.

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[5] 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 significantly (Table 3), whereas the head teacher actions to improve pupil attendance has decreased significantly. The proportion of schools with two or more activities to improve access for disadvantaged children in their SDP more than doubled from CS2 to CS3. Overall, across all schools in the six states we do not find a significant difference between numbers of schools which met the inclusiveness standard in CS3 vis-à-vis CS2. We also find that significantly more schools failed the partial inclusiveness standard in CS3, as compared to CS2.

---

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

Table 3:   School inclusiveness in CS1, CS2 and CS3

	2012 (CS1)	2014 (CS2)	2016 (CS3)	2012–16	2014–16
(1) Three or more actions on learner attendance (%)	57.9	28.7	26.5	-31.4*	-2.2
-- Number of actions on learner attendance	2.7	2.1	2	-0.8*	-0.1
(2) Two or more activities in SDP on access (%)	5.4	11.9	23	+17.7*	+11.2*
-- Number of activities on access	0.2	0.4	0.9	+0.7*	+0.4*
(3) >50% of teachers use two or more assessment methods (%)	70.7	62.3	55.1	-15.6*	-7.2
(4) >50% of teachers spatially inclusive and >50% are gender inclusive (%)	33.4	23.4	21.5	-11.9*	-1.9
Number of inclusiveness criteria fulfilled (out of four)	1.7	1.3	1.3	-0.4*	+0.0
Inclusiveness score (%)	72.2	63	83.8	+11.5*	+20.8*
School fully met standard (three to four criteria) (%)	18.8	10.5	11.4	-7.4*	+0.9
School partially met standard (two to four criteria) (%)	60.4	43	34.5	-25.9*	-8.5*

Note: * indicates significant results.

Within CS3, schools with more ESSPIN intervention are more inclusive than other schools (Table 4). Schools which received ESSPIN Output 4 intervention are significantly more inclusive than 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 the inclusiveness standards. On average, non-ESSPIN schools just met one criterion for school inclusiveness, as compared to ESSPIN schools, which met between 1.4 and 1.6 criteria, depending on whether the intervention was received prior to the first Composite Survey, or after it. Overall, 13.5% and 19.1% % of ESSPIN pre-CS1 and post-CS1 schools (respectively), and only 7.8% of non-ESSPIN schools, fully met the school inclusiveness standard. Teachers in post-CS1 schools also used more assessment methods and involved children from more parts of the classroom in lessons, whereas teachers in both pre-CS1 and post-CS1 schools had more activities in the SDP for disadvantaged children. Amongst the ESSPIN schools, the results also indicate that post-CS1 schools have more positive and significant results as compared to pre-CS1 schools.

These findings are consistent with the findings of the Inclusive Education Review of ESSPIN schools (Pinnock, 2016) that found that ESSPIN has instigated a wide range of activities in all the states to boost inclusive education. These activities included awareness raising campaigns at the state and LGEA levels for enrolling disabled and vulnerable children in schools, enrolment drives targeting ethnic, gender and disabled minorities, and even small-scale funding efforts for schools to support disabled learners and bring mainstream schools closer together (Pinnock, 2016). These activities were carried out by the SBMCs. The report also found that ESSPIN played a key role in promoting many of these changes and activities, and complemented ESSPIN Output 4 interventions. These could be indicative reasons for the better performance of ESSPIN schools, vis-à-vis non-ESSPIN schools.

Table 4:   School inclusiveness in CS3, by Output 4 intervention group

School inclusiveness criteria	Control	Pre-CS1		Post-CS1	
Inclusiveness criteria					
(1) Three or more actions to improve attendance (%)	28.3	16.1		24.6	
(2) Two or more activities in SDP to improve access for disadvantaged children (%)	17.4	33.8	+	33.8	+
(3) >50% of teachers use two or more assessment methods (%)	44.6	70.8	+	76.2	+
(4) >50% of teachers spatially inclusive and >50% are gender inclusive (%)	19.4	15.4		27.8	
Overall inclusiveness standard					
Number of inclusiveness criteria fulfilled (out of four)	1.1	1.4	+	1.6	+
School partially met inclusiveness standard (two to four criteria out of four)	25.8	38.7		53.6	+
School fully met inclusiveness standard (three to four criteria out of four) (%)	7.8	13.5		19.1	+

4.2   Head teachers' actions to improve pupil attendance

In all three rounds of the survey, head teachers in sampled schools were asked questions on actions taken, if any, to improve pupil attendance in their schools. We find that there is a weak, significant difference between the actions taken by head teachers to improve pupil attendance in CS3 as compared to CS2. There is a significant decline in keeping up-to-date registers to improve pupil attendance in CS3 as compared to CS2, whereas there is a significant increase in involving SBMCs in finding reasons for non-attendance, and implementing suggested solutions for non-attendance. This is consistent with the findings of Usman (2016), who highlighted an increase in SBMCs' involvement in promoting timely pupil and teacher attendance through an increase in SBMCs' monitoring of attendance, and moral support given to pupils for good attendance and behaviour.

Table 5: **Head teachers' actions to improve pupil attendance in CS1, CS2 and CS3 schools**

% of head teachers who took the following actions	CS1	CS2	CS3	Significant diff from CS2 to CS3
Keep up-to-date registers	45.3	44.3	21.4	-
Involve SBMC in finding reasons for non-attendance	45.0	40.4	58.1	+
Discuss with teachers, pupils or parents about reasons for non-attendance	74.3	72.6	74.9	
Implement suggested solutions for non-attendance	27	7.7	13.4	+
Address issues such as bullying and punishment	22.2	3.3	1.4	
Improve quality of teaching and learning	29.8	6.7	4.4	
Average number of actions taken	2.75	2.06	2.0	

Focusing on CS3 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 ESSPIN schools reported keeping up-to-date registers as a means to curb pupil absenteeism, and implementing suggested solutions to address the issue of non-attendance. Even though there was a significant increase over time in the involvement of SBMCs for finding reasons for non-attendance (reported in Table 5), there is no significant difference between control schools and ESSPIN schools on this measure. This could be viewed as a positive sign, implying that it is not just the SBMCs in ESSPIN schools, but also the SBMCs in control schools that were engaging in issues concerning pupil attendance. Overall, for the majority of the actions listed, the percentage of head teachers using these as means to improve pupil attendance was not significantly different between schools receiving ESSPIN Output 4 interventions and control schools.

Table 6: **Head teachers' actions to improve pupil attendance in CS3 schools, by Output 4 intervention groups**

% of head teachers who took the following actions	Control	Pre-CS1	Post-CS1	
Keep up-to-date registers	24.1	14.6	16.4	
Involve SBMC in finding reasons for non-attendance	57.1	69.4	58.2	
Discuss with teachers, pupils or parents about reasons for non-attendance	74.0	80.1	76.0	
Implement suggested solutions for non-attendance	15.2.5	7.1	10.2	
Address issues such as bullying and punishment	0.9	1.0	2.8	
Improve quality of teaching and learning	4.9	3.6	3.5	
Average number of actions taken	2.0	2.0	1.9	

We further analyse these results disaggregated by head teachers' gender in Section 8.4.

4.3   Spatial and gender inclusiveness

All three rounds of the Composite Surveys observed sampled teachers' classroom practices, including spatial and gender inclusiveness. In particular, spatial inclusiveness was observed using the classroom observation instrument. Data collectors were trained to divide the classroom virtually into six zones and to note the number of instances when the teacher engaged with a pupil or pupils

from each of the six zones. As such, by spatial inclusiveness we are interested in whether the teacher engaged with at least one pupil from four different areas of the classroom during a lesson.

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 to their presence in the classroom, within a 10% margin. For example, if the class contains 50% girls then teachers who engage with girls in between 60% and 40% of total engagements 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.[6]

Using these definitions and indicators, we tabulate results on spatial and gender inclusiveness below. Table 7 shows that there was no significant difference for teachers sampled in CS3 schools, on average, on spatial inclusion than teachers sampled in CS2. In terms of gender inclusiveness, no significant differences were found between teachers in CS2 and CS3 schools, except for involving girls and boys proportionately (within 10%) in class. In CS3 schools, only 45.7% of the teachers observed involved boys and girls proportionately (within 10%), as compared to 52% in CS2 schools – a statistically significant difference.

Table 7:   Spatial and gender inclusiveness by teachers in CS1, CS2 and CS3 schools

Indicators (%)	CS1	CS2	CS3	Significant diff. between CS2 and CS3
Participation per minute in class	27.9	33.8	31.9	
Involves boys/girls proportionately (within 10%)	47.6	52.1	45.9	-
Gender equity score (0=completely unequal, 100=perfectly equal)	81.6	82.5	81.9	
Involves pupils from four or more areas of the class	75.2	59.4	62.4	
Number of zones participating in lessons (#)	4.27	3.76	3.89	

Do teachers in ESSPIN Output 4 schools demonstrate more gender and spatial inclusiveness than their counterparts in control schools? As Table 8 shows, teachers who are in ESSPIN schools perform significantly better than those in control schools, i.e. non-ESSPIN schools, in terms of both the spatial inclusiveness and gender inclusiveness criteria.[7] This is consistent with the findings of Pinnock (2016) who highlights the key role played by ESSPIN in promoting greater inclusiveness, with a focus on vulnerable, disabled, gender and ethnic minorities. Some of the gender-friendly measures included investment in more accessible infrastructure, such as gender accessible latrines, summer camps and sports for girls in Jigawa and Kano to increase awareness and participation, and a more over-arching framework for SBMCs to promote inclusive practices nationally. The report found that ESSPIN efforts have had a significant impact on promoting inclusive education across all six states, and on providing impetus to community involvement. It has also had a significant impact in boosting the uptake of measures aimed at targeting girls' education. Since CS2, ESSPIN has also developed new strategies to target girls and other vulnerable groups. However, for the gender equity score, the difference between control schools

---

[6] 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.

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

and ESSPIN schools is only significant when we compare the schools that started receiving Output 4 interventions recently (the post-CS1 schools) to the control schools. On the whole, we also find that ESSPIN schools that started receiving Output 4 interventions after CS1, i.e. the post CS1 ESSPIN schools, have more significant and positive results as compared to pre-CS1 ESSPIN schools.

Table 8: **Spatial and gender inclusiveness by teachers in CS3 schools, by Output 4 intervention groups**

Indicators (%)	Control schools	Pre-CS1 schools		Post-CS1 schools	
Participation per minute in class	28.3	30.6		37.1	+
Involves boys/girls proportionately (within 10%)	42.6	42.4		51.6	+
Gender equity score (0=completely unequal, 100=perfectly equal)	79.3	82.8	+	85.8	+
Involves pupils from four or more areas of the class	55.5	67.2	+	71.0	+
Number of zones participating in lessons	3.7	4.0	+	4.2 4	+

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 CS2 and CS3. SBMCs in ESSPIN schools were also typically much more functional than those that had not received ESSPIN intervention. Amongst ESSPIN schools, we also find that schools that started receiving Output 4 interventions after CS1, i.e. post-CS1 ESSPIN schools, have relatively higher and more positive results than pre-CS1 ESSPIN schools. There was a significant change over time in the participation of women and children in SBMCs. The increase in participation of women in SBMCs was driven by an increase in women's attendance, issues raised by women in SBMC meetings, and actions taken on issues raised by women; whereas the increase in children's inclusiveness in SBMCs was driven by an increase in children's attendance, and actions taken on issues raised by 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[8]:
1) Two or more SBMC meetings have taken place since the start of 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) SBMC women's committee exists (written or physical evidence)
7) SBMC children's committee exists (written or physical evidence)
8) SBMC contributed resources for the school (written or physical evidence)
9) SBMC chair has 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 has improved between CS2 and CS3, although the difference is not significant for every indicator (Table 9). In total, across the 6 states, we estimate that 5100 more schools have functional SBMCs in 2016 than in 2012. As discussed above, these results are for all schools in the six states, covering both programme and control schools; they thus provide an overview of the state of affairs rather than of programme impact. The average school met three of the nine criteria in 2014, but in 2016 met over five 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 record-keeping of the SBMC more than the particular aspects of functionality they aim to measure. There has been no significant decline in any of the criteria between 2014 and 2016.

---

[8] 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.

Examining individual criteria, SBMCs have particularly improved in terms of networking with other CBOs, institutions and/or SBMCs, addressing exclusion, number of meetings, awareness raising and interacting with LGEA. This is consistent with Usman (2016), who conducted an in-depth review of the resource mobilisation and disbursement efforts of the SBMCs and found that a vast majority of these resources were used to improve the school and access to education. SBMCs also undertook numerous activities to address the exclusion of children from school, including training and mentoring sessions for schools on more inclusive education practices, such as training teachers to support children with disabilities, supporting the poorest children to access schools, addressing issues of gender discrimination, enrolment drives, and conducting out of school surveys to identify the children that are out of school (Pinnock, 2016). The positive action of the SBMCs in addressing exclusion is in stark contrast to the finding that SBMCs do not lead to increases in enrolment or decreases in drop-out rates (International Initiative for Impact Evaluation, 2016).

From the CS3 results it is clear that there has been a significant and positive improvement in criteria 3, i.e. SBMCs' actions to address exclusion, but no significant change in the criteria specifically related to exclusion, i.e. actions taken for commonly excluded groups, and issues raised regarding children's exclusion. In order to explain this discrepancy, it is important to understand the questions on which these indicators are based. Criteria 3, i.e. SBMC addresses exclusion, is based on a question that explicitly asks if the SBMCs took any actions to address the attendance and prevent drop-out. On the other hand, the criteria on actions taken for commonly excluded groups is based on a question that requires SBMCs to assess the actions that they have taken to support commonly excluded groups, whereas raising issues relating to children's exclusion is based on a question that requires SBMCs to evaluate the issues raised that pertain to children's exclusion. As compared to the question on attendance and drop-out, which is easily understood, easy to assess and record (especially since there are follow-up questions on what measures were taken by the SBMCs to address attendance and prevent drop-out), the other two questions relating to commonly excluded groups, and issues that are related to children's exclusion are more complex, and could be subject to misinterpretation. For instance, 'commonly excluded groups' may be differently understood by different interviewees, and even though the interviewer is prompted to explain what commonly excluded groups could refer to, the understanding of this completely depends on the communication between the interviewer and the interviewee. The same is true for issues relating to children's exclusion. Also, the latter two criteria rely heavily on written, oral or photographic evidence, which may be difficult for the SBMCs to produce, given the wide scope of these questions. Thus, some part of this discrepancy can be explained by measurement error. Other criteria measuring SBMCs actions on inclusion and exclusion are explained in the table below.

Box 3: Asking SBMCs about inclusion and exclusion

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

Criterion	Question asked (with data collector instructions in blue)	Criterion met if…

(2) Conducted awareness raising	Did the SBMC do anything to raise awareness about the value of education for all boys and girls in the community in the current school year?	Respondent answers yes and can present oral or written evidence
(3) Addressed exclusion	Did the SBMC do anything to address issues which prevent children from attending school or which cause drop-out in the current school year?	Respondent answers yes and can present oral or written evidence
(A1) Took action for commonly excluded groups	Did the SBMC do anything to **support commonly excluded groups** in the **current school year**? You can explain that commonly excluded groups could be orphans, nomadic children, girls, children with disability, ethnic or religious minorities, etc.	Respondent answers yes and can present oral or written evidence
(A2) Raised issues of children's exclusion	Did the SBMC **raise issues of children's exclusion** from school in the community, with the LGEA, or with the state government, in the **current school year**?	Respondent answers yes and can present oral or written evidence
(A10) Raised cash to support vulnerable children	Did the SBMC mobilise any cash to support vulnerable children in the current school year?	Respondent answers yes (no evidence required)
(A11) Monitored drop-out or non-attendance (A12) Communicated with school or community about drop-out (A13) Number of actions taken to address non-attendance	**What actions were taken** to address issues which prevent children from attending school or which cause drop-out in the **current school year**? Do not prompt. This is a multiple response question –- **SELECT ALL THAT APPLY** • Monitoring drop-out • Monitoring non-attendance • Communicating with school about drop-out • Communicating with community about drop-out • Other (specify) • Don't know / refused	Respondent answers yes to a previous question (asking whether any action was taken to address these issues) and then provides this information in the follow-up question on what type of action and how many actions were taken. No specific evidence is required

The analysis also indicates that there is no significant difference between the resources contributed to the schools by SBMCs between CS2 and CS3. This supports the findings of Usman

(2016), who finds that the official reports of resources raised and contributed are relatively rough estimates and may not accurately account for the value of material and in-kind resources or other donations received from outside communities (Usman, 2016). The validation study on resource mobilisation by Usman (2016) found that the resources raised by SBMCs are considerably higher than the amount reported in the state monitoring reports.

Two additional criteria related to actions taken by SBMCs to address exclusion 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. There has been no significant change for either of these criteria.

Table 9: SBMC functionality in CS1, CS2 and CS3

	2012 (CS1)	2014 (CS2)	2016 (CS3)	2012–16	2014–16
(1) Two or more meetings this school year (%)	28.7	27.1	52.1	+23.3*	+25.0*
(2) Conducted awareness raising (%)	35.3	47.5	70.4	+35.1*	+22.9*
(3) Addressed exclusion (%)	26.7	40.1	67.5	+40.8*	+27.4*
(4) Networked with CBOs/institutions/other SBMCs (%)	15	55.6	94	+78.9*	+38.4*
(5) Interacted with LGEA (%)	19.7	21.1	38.4	+18.7*	+17.3*
(6) Has a women's committee (%)	13.1	26.6	41.4	+28.3*	+14.7*
(7) Has a children's committee (%)	19	21	35.8	+16.8*	+14.8*
(8) Contributed resources for school (%)	39	54.5	61.2	+22.2*	+6.6
(9) Chair visited school three or more times (%)	25.2	14.8	24.7	-0.5	+9.9*
Number of SBMC functionality criteria met (out of nine)	2.3	3.3	4.6	+2.4*	+1.4*
School meets standard for functioning SBMC (%)	21.7	30.9	56.1	+34.4*	+25.1*
Inclusion and drop-out					
(A1) Took action for commonly excluded groups (%)	13.9	23.8	18.7	+4.8	-5.1
(A2) Raised issues of children's exclusion (%)	4.8	19.3	14.2	+9.4*	-5.1

Examining differences between ESSPIN Output 4 schools in CS3 vis-à-vis the control schools (Table 10), there were significant differences in the expected direction across most of the criteria. The average ESSPIN Output 4 school met around six of the nine criteria, while control schools met around 4.8 – a statistically significant difference. Overall, 60.9% and 70.7% of pre- and post-CS1 schools met the SBMC functionality standard, compared to 49.7% control schools, with the difference for post-CS1 schools being statistically significant. However, there was no significant difference between control schools and ESSPIN Output 4 schools for addressing exclusion, interacting with LGEA, taking action for commonly excluded groups and raising issues of children's exclusion. We also find that post-CS1 schools have relatively more significant and positive results as compared with pre-CS1 schools, which can be attributed to the fact that results are found to be the greatest in the first year of SBMC activation (Usman, 2016). This has important implications in regard to continued support of the government, both at the local and national level, for SBMCs. (Usman, 2016).

Table 10: SBMC functionality in CS3, by Output 4 intervention groups

SBMC functionality criteria	Output 4 control	Pre-CS1		Post-CS1	
(1) two or more meetings this school year (%)	43.1	55.8		72.1	+
(2) conducted awareness raising (%)	67.1	81.3	+	75.9	
(3) addressed exclusion (%)	66.3	82.2	+	67.2	
(4) networked (%)	93.9	97.9	+	93.1	
(5) interacted with LGEA (%)	37.5	30.2		42.1	
(6) has women's committee (%)	29.4	60.8	+	65.1	+
(7) has children's committee (%)	21.9	53.7	+	64.2	+
(8) contributed resources for school (%)	60.7	73.5		59.5	
(9) chair visited school three or more times (%)	21.1	30.4		31.7	
Standard S: functioning SBMC	49.4	61.0		70.7	+
Number of SBMC functionality criteria met (out of nine)	4.2	5.5	+	5.5	+
Additional criteria					
Action for commonly excluded groups	18.6	25.1		17.4	
Raised issue of children's exclusion	13.9	21.6		13.3	

5.2 Participation of women in SBMCs

As in the previous report on CS1 and CS2, 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 4).

Box 4: 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) Female member of SBMC raised at least one issue at SBMC meetings (written evidence or oral evidence from 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)

4) At least one SBMC women's committee meeting took place[9]

There was a significant improvement between CS2 and CS3 for all the criteria for women's inclusiveness in SBMCs, excluding women's committee meetings. This improvement was especially high for issues raised by women in SBMC meetings and action taken on issues raised by women. Overall, there was a statistically significant improvement in the % of schools that meet the women's inclusiveness standard, and the number of criteria met by average schools. In 2014, an average school met less than one criteria for women's inclusiveness, whereas in 2016, an average school meets almost two criteria. This is consistent with Pinnock (2016), who finds that SBMC women's committees have been strengthened since 2014, when the second Composite Survey was conducted. The report notes that in the previous review of the women's committees in 2014, women exhibited varying levels of confidence, whereas all the women's committees in 2016

---

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

exhibited 'strong confidence and awareness of a clear mission'. The report also notes that members of SBMCs' women's committees frequently checked teachers' attendance and also lobbied to address teacher absenteeism. The dynamic role of the women's committees was also exhibited in more conservative areas like Kano (Pinnock, 2016).

Table 11:   SBMC's women's inclusiveness in CS1, CS2 and CS3

Participation of women in SBMC	2012 (CS1)	2014 (CS2)	2016 (CS3)	Change in average over time	
				2012–16	2014–16
(1) At least one woman attended two or more meetings (%)	19.5	17.4	29.6	+10.1*	+12.2*
(2) Female member raised an issue (%)	26.8	31.9	54.9	+28.1*	+23.1*
(3) Issue raised by female member led to action (%)	28.3	14.5	31.8	+3.5	+17.3*
(4) Women's committee met (%)	7.8	27.1	31.6	+23.8*	+4.5
No. criteria met (out of four)	0.6	0.9	1.3	+0.6*	+0.4*
Meets standard (three out of four criteria) (%)	15.4	15.5	23.2	+7.8	+7.7*

Examining differences between ESSPIN Output 4 schools in CS2 (Table 12), there were significant differences for both types of Output 4 schools in the expected direction in all four criteria and in the overall standard. ESSPIN schools, both the ones receiving the Output 4 interventions prior to CS1, as well as those receiving the interventions after CS1, performed significantly better than the control schools, i.e. the non-ESSPIN schools, for all the indicators for women's inclusiveness in SBMCs. The average ESSPIN Output 4 school met over 2.5 of the four criteria, while control schools met around 1.5 on average. Overall, 11.2% of control schools are meeting the women's inclusiveness standard, as compared to 42.3% and 47.2% of pre-and post-CS1 ESSPIN schools, respectively. Although both pre-CS1 and post-CS1 ESSPIN schools have positive and significant results as compared to control schools, post-CS1 ESSPIN schools have relatively higher and better results.

Table 12:   SBMC women's inclusiveness in CS3, 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 (%)	18.1	42.0	+	53.6	+
(2) female member raised an issue (%)	44.7	71.7	+	66.1	+
(3) issue raised by female member led to action (%)	23.7	46.4	+	47.6	+
(4) women's committee met (%)	20.6	51.8	+	53.0	+
Number of criteria met	0.8	1.9	+	2.1	+
Meets standard (three out of four criteria)	11.2	42.5	+	47.2	+

5.3   Participation of children in SBMCs

We measure SBMC children's inclusiveness using four criteria (Box 5).

Box 5: 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)   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)

4)   At least one SBMC children's committee meeting took place and committee has a trained facilitator[10]

There was a significant increase overall between CS2 and CS3 in the number of meetings attended by children and action taken on issues raised by children. This was also accompanied by an increase in the proportion of schools that meet the children's inclusiveness standard, which increased from 6.2% in CS2 to 9.7% in CS3. The number of criteria met by an average school also showed a small, but significant increase from just over 0.5 to almost 1 criterion. As discussed above, these results are for all schools in the six states, covering both programme and control schools; they thus provide an overview of the state of affairs, rather than of programme impact. The significant and positive changes between CS2 and CS3 for SBMC children's committees are consistent with the fact that children's committees had not initially received direct support in the form of mentoring visits from CSOs and SMOs, but post-2014, i.e. post-CS2, ESSPIN organised direct training for children's committees (Pinnock, 2016). However, there were concerns about the issues raised by the children's committees, which included better toilets, sports equipment and music equipment, amongst others, but did not address general issues relating to exclusion, as well as issues relevant for disabled and other vulnerable sections (Pinnock, 2016).

While evaluating the benefit of involving children in SBMCs, it is also important to note the issues brought up by children in the committee meetings. For instance, the children in the committees did have good knowledge of exclusion issues that affected children in schools, even though they were not discussing this or raising these issues in the committee (Pinnock, 2016).

---

[10] 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 failure to keep minutes does not mean the committee is not functioning.

Table 13:   SBMC children's inclusiveness in CS1, CS2 and CS3

Participation of children in SBMC	2012 (CS1)	2014 (CS2)	2016 (CS3)	Change in average over time	
				2012–16	2014–16
(1) At least one child attended two or more meetings (%)	11.8	8.8	19.6	+7.8*	+10.8*
(2) A child raised an issue (%)	13.6	20.6	22.1	+8.5*	+1.5
(3) Issue raised by child led to action (%)	11.5	7.3	22.4	+10.9*	+15.1*
(4) Children's committee met (%)	2.4	14.3	19.2	+16.9*	+4.9
No. criteria met (out of four)	0.3	0.5	0.8	+0.5*	+0.2*
Meets standard (three out of four criteria) (%)	5.7	6.2	9.7	+4.0	+3.5*

As with women's inclusiveness, there are large, positive and significant differences between ESSPIN and non-ESSPIN schools, with the schools receiving ESSPIN Output 4 interventions having SBMCs that are more inclusive of children as compared to the control schools. Disaggregating CS3 schools by Output 4 and control schools (Table 14), we find large positive and significant differences on all criteria for SBMC children's inclusiveness in the expected direction. Overall, 25.2% and 19% of pre- and post-CS1 schools met the logframe standard for children's inclusiveness, while only 4.4% control schools did so.

Table 14:   SBMC children's inclusiveness in CS3, between Output 4 intervention groups

SBMC children's inclusiveness criteria	Control	Pre-CS1		Post-CS1	
(1) child attended two or more meetings (%)	16.5	32.7	+	23.9	
(2) child raised an issue (%)	14.0	39.2	+	35.8	+
(3) issue raised by child led to action (%)	19.0	30.1		28.7	
(4) children's committee met and it has a trained facilitator (%)	9.0	40.8	+	38.6	+
Number of criteria met	0.5	1.3		1.2	+
Meets standard (three out of four criteria) (%)	4.4	25.2	+	19.0	+

PART B: DIFFERENCES IN EDUCATION OUTCOMES BY GENDER AND BACKGROUND

6   Introduction to Part B

Do boys who are in school learn more than girls? Are students from poorer families, over-age 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 bigger 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 Grades 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, and so on) and apply statistical tests to assess whether any differences in our estimates are likely to have occurred by chance or if they 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 Output 3 interventions aiming to strengthen 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 quality instruction, and 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 improving quality gains.[11] 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.

In this section, we also disaggregate some of the results by state. In order to better understand the state-wise breakdown of the results, it is also important to briefly focus on the contextual factors in all the states. For instance, in Lagos, the political situation is stable, and there are also no major issues of conflict. There are many other donor and privately-funded education programmes in Lagos, which are complementary to ESSPIN, and, hence, are seen as having a positive effect on ESSPIN schools. On the other hand, in Kano, insecurity and violence remains a major concern, although the state has witnessed a decline in the number of incidents of political violence. In Kaduna, political changes have led to new programmes on free feeding and school uniforms, which has led to an overcrowding of classrooms, and a consequent decline in teachers' control and students' attention, which may have a bearing on the pupils' performance in class. There were also issues around teachers' payments, with some teachers not having been paid for eight months. This is likely to have a negative impact on teachers' motivation.

In Enugu there is clear support for ESSPIN, as the new education commissioner is in favour of supporting education reform and improving the quality of schools. However, teacher recruitment is an issue in Enugu, and there are a disproportionate number of teachers in the urban areas as compared to the rural areas. Invasion of villages by armed herdsmen is also a growing issue in the region. Political violence and insecurity remains a major threat in Jigawa, although there has been a slight decrease in the number of incidents of political violence in the state between 2014 and 2015. There has been a large increase in the enrolment of students in Jigawa, but no corresponding increase in teachers. With teacher retiring, schools are faced with staffing issues, and the pressure on the remaining teachers in the system has increased. In Kwara there has been an increase in the enrolment of students in public schools, as well as a transfer of children from

---

[11] Scale-up of the SIP is discussed in more detail in the overall CS2 report (Cameron, 2015).

public to private schools due to improved infrastructure and better SBMCs. However, this has led to an increasing pressure on existing teachers. Also, teachers in the state are shifting from rural schools to urban schools, which has led to a shortage of teachers in rural areas in the state.

For ease of reading the following sections we divide the schools into three categories – minimum intervention, medium intervention, and maximum intervention – on the basis of the number of years of Output 3 intervention under ESSPIN. *Minimum refers to those schools that have received less than or equal to one year of Output 3 intervention, medium refers to the schools that have received two to three years of intervention under Output 3, and maximum refers to those schools that have received four to five years of intervention under Output 3.*

7   Pupil learning outcomes

7.1   Gender differences

The mean item response theory (IRT) test scores disaggregated by the gender of the pupils indicates that boys perform significantly better than girls on all tests, except for the Grade 2 numeracy test. The difference between the mean test scores between boys and girls increases with their grade. For instance, the difference between boys and girls on the mean test scores for the literacy test in Grade 2 is 9.9 points, whereas that for the Grade 4 literacy test it is 18 points. The same trend is observed for the numeracy scores, though the differences are only significant for the Grade 4 numeracy test.

Table 15:   Gender disaggregated pupil learning outcomes in CS3

Mean test score %	Boys	Girls	Significant diff.
N2	442.2	447.3	
L2	451.6	441.8	Boys
N4	469.2	450.7	Boys
L4	456.3	438.3	Boys

In order to test whether there was a difference between the performance of girls and boys in schools that received differing amounts of intervention under Output 3 we disaggregate the test scores on the basis of gender and the amount of intervention, i.e. minimum, medium or maximum. We find that for schools that received minimum interventions under Output 3 there is a significant difference between the test scores of boys and girls, with boys performing significantly better than girls on all tests except the numeracy tests for Grade 2. There is no significant difference between the performance of girls and boys for schools that received medium interventions. For schools that received maximum interventions, there is no clear trend in the performance of boys and girls, with boys perform significantly better than girls on the literacy test in Grade 2, but girls performing significantly better than boys in Grade 4. There is no significant difference for numeracy tests.

Table 16:   Gender disaggregated pupil learning outcomes in CS3 by intervention groups

Mean test score (%)	Minimum intervention (zero to one years)			Medium intervention (two to three years)			Maximum intervention (four to five years)		
	Boys	Girls	Sig. diff	Boys	Girls	Sig. diff	Boys	Girls	Sig. diff
L2	438.2	424.5	Boys	467.8	475.6		504.4	488.0	Boys
L4	437.3	420.3	Boys	481.4	463.3		483.7	506.1	Girls
N2	430.2	436.8		455.6	463.4		496.3	489.5	
N4	456.2	424.8	Boys	484.9	485.5		513.2	511.2	

In order to determine if there are any state-level variations in the performance of girls and boys, we disaggregate the mean test scores by gender and state (Table 17). There is no overall trend that emerges. In Kano, boys perform significantly better than girls on all tests, except the Grade 2 numeracy test, whereas in Lagos, girls perform significantly better than the boys on the Grade 4 literacy tests and the Grade 2 numeracy test. For the other states, no consistent and significant trend emerges.

Table 17:   Gender and state disaggregated pupil learning outcomes in CS3

	Enugu			Jigawa			Kaduna			Kano			Kwara			Lagos		
	B	G	Sig. diff	B	G	Sig. diff	B	G	Sig diff	B	G	Sig. diff	B	G	Sig diff	B	G	Sig diff
L2	578.9	581.4		427.1	421.3		459.9	441.6		428.7	418.3	Boys	484.6	475.5		621.2	629.5	
L4	551.1	572.1	Girls	399.1	401.4		453.6	441.9		444.1	407.7	Boys	469.6	480.6		586.7	608.7	Girls
N2	537.9	547.2		401.0	436.6	Girls	438.2	475.1		428.8	417.7		493.5	472.9	Boys	596.1	628.5	Girls
N4	528.8	524.4		433.8	422.7		469.7	444.3		454.8	423.3	Boys	499.7	501.4		607.4	603.4	

7.2   Wealth

Did poorer pupils perform differently to 'richer' pupils on the CS3 pupil tests? In CS3, data on pupil household wealth were collected using a short asset questionnaire administered at the end of the Grade 4 literacy and numeracy tests. Data collectors showed Grade 4 pupils pictures of common household assets (given in Box 6), and asked the pupil if his/her family had that asset at home.

Box 6: Question in CS3 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. 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.

In order to determine if there is a wealth gap in pupil test scores, we run a statistical model, regressing the mean literacy and numeracy scores in Grade 4 on three covariates: an indicator of wealth, an indicator of the extent to which the child's school has been exposed to ESSPIN Output 3[12] interventions, and an interaction term between the two variables. We find that wealth has a positive and significant impact on both the literacy and numeracy scores of the pupils, whereas the amount of exposure to ESSPIN interventions only has a positive and significant impact on literacy, but not on numeracy scores. However, we are unable to establish causation, because selection of ESSPIN schools is not random and we do not have data on the wealth of children at different points in time. The interaction term between these two variables is not significant for either literacy or numeracy, indicating that the impact of wealth on test scores does not change with the amount of ESSPIN intervention, and vice-versa.

The following figure depicts the average Grade 4 literacy scores against the wealth quintiles of the pupils.

---

[12] A significant negative coefficient on the interaction term indicates that wealth inequality in learning outcomes is smaller in schools with more ESSPIN intervention.

Figure 3:   Grade 4 literacy scores by wealth quintiles
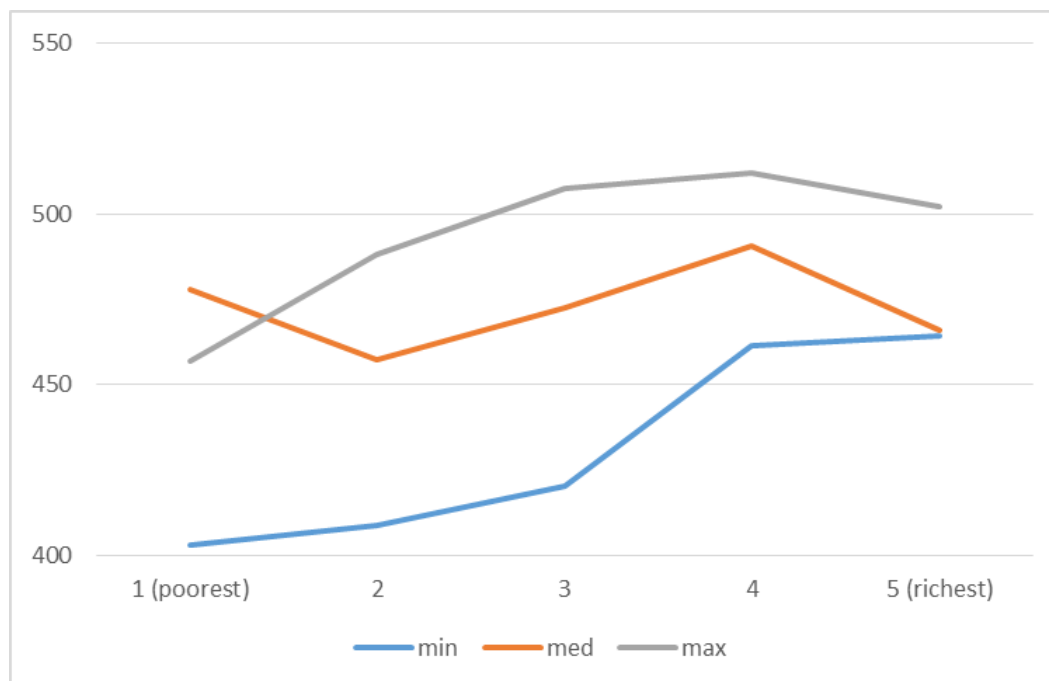
Source: Author's own calculation

It can be seen from the table that for schools receiving minimum intervention under Output 3 there is a huge difference between the poorest students and the richest students. This difference is much less great for schools that received maximum intervention. No clear trend is observed for the schools receiving the medium amount of intervention.

The following figure shows the average Grade 4 numeracy test scores by wealth quintiles.

Figure 4:   Grade 4 numeracy scores by wealth quintiles
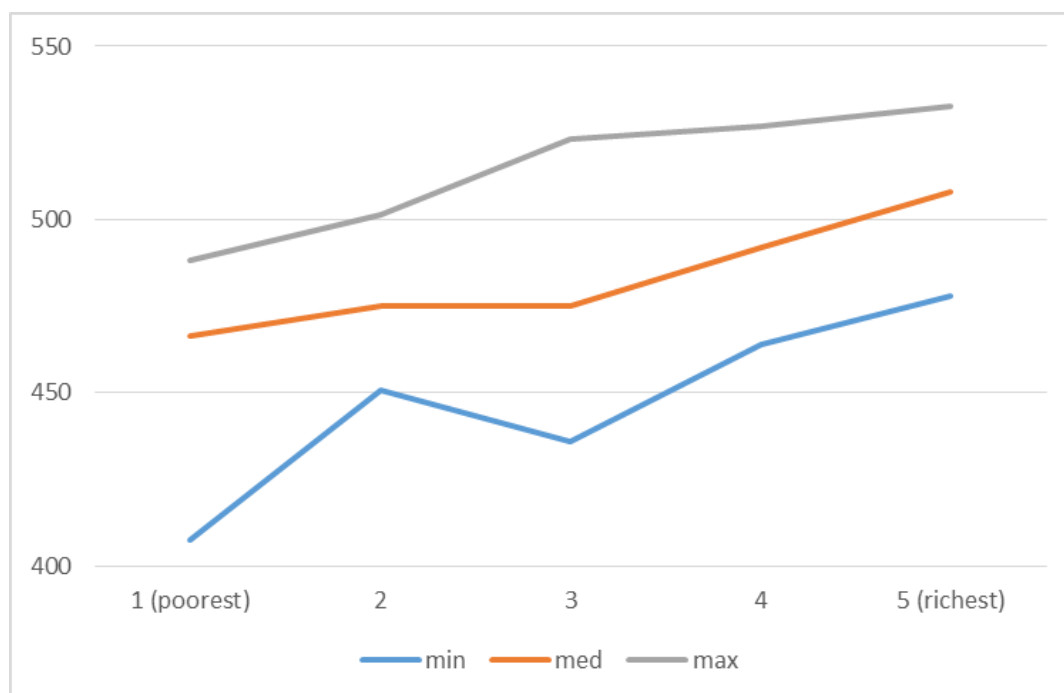
Source: Author's own calculation

It can be seen from this graph that there is a substantial and marked difference between the poorest and the richest pupils in the schools receiving minimum intervention. This difference persists, but is relatively smaller, in schools receiving medium and maximum intervention. In order to better understand the pupils in the poorest quintile, we further disaggregate the results of students in the poorest quintile by whether the pupils in Grade 4 meet the numeracy and literacy standards of Grade 4 or whether they are performing at a lower standard, i.e. at the standard of Grade 2 pupils, by gender, and by the amount of intervention received by the respective schools under ESSPIN Output 3. This is shown in the following table.

Table 18:   Pupils in the poorest quintile by performance level (Grade 2/Grade 4), gender and ESSPIN Output 3 intervention

Pupils in the poorest wealth quintile	All (%)	Male (%)	Female (%)	Minimum Output 3 intervention (%)	Medium /maximum Output 3 intervention (%)
Grade 4 Literacy standard	12.0	17.4	4.8	1.2	33.2
Grade 4 numeracy standard	4.7	2.9	7.1	0.8	13.1
Grade 2 literacy standard	19.3	24.3	12.6	4.6	48.0
Grade 2 numeracy standard	22.3	25.6	17.9	17.0	33.8

The above table shows the percent of pupils in Grade 4 that meet either the Grade 4 or Grade 2 literacy and numeracy standards overall, by gender, and by the amount of ESSPIN Output 3 intervention received. It can be seen from the above table that there are wide disparities even amongst the pupils in the poorest quintile. Only 12% and 4.7% of the pupils in Grade 4 meet the literacy and numeracy standards applicable to Grade 4. Of these, 33.2% of the pupils are from schools that have received medium or maximum intervention under ESSPIN Output 3 and only 1.2% from schools receiving minimum Output 3 intervention. 17.4% of the male pupils meet the Grade 4 literacy standards, as compared to only 4.8% of female pupils. For Grade 4 numeracy standards, the opposite is true, with 7.1% of female pupils meeting the Grade 4 numeracy standard, as compared to 2.9% of male pupils. For the Grade 4 pupils who only meet the Grade 2 literacy and numeracy standards, we observe the same trends for the amount of ESSPIN Output 3 intervention received, with the percent of pupils from schools receiving medium or maximum Output 3 intervention far exceeding the percent of pupils coming from schools with minimum Output 3 interventions. The percent of male pupils amongst the poorest quintile who meet the Grade 2 literacy and numeracy standards exceeds the percent of female pupils. Another over-arching indication of these results is that a greater percent of pupils from the poorest wealth quintile in Grade 4 meet the Grade 2 literacy and numeracy standards, as opposed to the Grade 4 literacy and numeracy standards.

7.3   Speaking a minority language

In CS3, sampled pupils in both Grades 2 and 4 were asked to name the main language they speak at home. Box 7 shows the question asked in the pupil tests and the options provided.

Box 7: Question in CS3 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 19). Lagos is more mixed than the other states, with 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 19:   Main language spoken by pupils at home in CS3, by state

Language (%)	Enugu	Jigawa	Kaduna	Kano	Kwara	Lagos
Benin / Edo	0	0	0	0	0	1.5
Ebira	0	0	0	0	0	0.2
English	6.1	0	2	0	2	20
Esan	0	0	0	0	0	0
Fulfulde	0	5	2	3	3	0
Hausa	0	94	73	94	1	2.5
Gbagyi	0	0	4	0	0	0
Gbari	0	0	0	0	0	0
Ibibio	0	0	0	0	0	0.5
Igbo	92	0	0	0	1	9
Kanuri	0	1	0	0	0	0
Nupe	0	0	0	0	9	0
Yoruba	0	0	0	0	66	59
Other/missing	2	0	17	0	16	7
Total	100	100	100	100	100	100

Note. Numbers in red indicate the main language in each state. The figures for Kaduna, Kano and Kwara do not add up to 100 as the remaining values (between 2% and 3%) were reported as 'don't know'.

In order to determine whether pupils who speak a minority language perform differently from those who speak a majority language, we disaggregate the test scores for literacy and numeracy by the language spoken by the student, i.e. whether it is a majority language or minority language. We find that for literacy tests, both for Grades 2 and 4, there is no significant difference between the pupils on the basis of the language spoken by them. However, for both the numeracy assessments we find that pupils speaking a majority language perform significantly better than those speaking a minority language. This is consistent with the general literature on the performance of pupils in mathematics when they are taught using their mother tongue and using cultural tools. For instance, Hafiz and Farik (2016) analyse the results of Grade 6 pupils in a school in Fiji, and find that primary school children learn mathematics more easily when it is taught in their mother tongue, and when their culture is embedded in the mathematical concepts and ideas (Hafiz and Farik, 2016).

Table 20:   Pupil test scores by language status in CS3

Mean test scores (%)	Minority	Majority	Significant diff.
L2	442.3	434.6	
N2	410.3	436.7	Majority
L4	422.9	434.6	
N4	430.1	448.8	Majority

Thus, speaking a minority language at home without support at school does not, on the basis of this evidence, appear to relate to a major impediment in literacy skills, but is linked to adverse outcomes in numeracy skills.

The results disaggregated by gender and language status are reported in the table below.

Table 21:   Pupil test scores by language status and gender

Mean test scores (%)	Minority			Majority		
	Boys	Girls	Sig. diff	Boys	Girls	Sig. diff
L2	442.2	448.6		442.6	428.7	Boys
N2	418.4	412.9		435.5	441.8	
L4	423.4	430.8		448.4	423.6	Boys
N4	434.5	436.5		464.0	435.0	Boys

It can be seen from the above results that among the majority language speakers, boys perform significantly better than girls on all tests except for the Grade 2 numeracy test, whereas there is no significant difference between the performance of boys and girls amongst minority language speakers.

7.4   Disability

7.4.1   Descriptive statistics from CS3

The questions to test the literacy and numeracy level of disabled children are detailed in the annex. In total, 51 pupils from the CS3 sample (i.e. less than 0.5%) were unable to perform one or more of the tasks as per the four screening questions (Table 22). 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 22:   Descriptive statistics on disabled pupils from CS3

	L2	L4	N2	N4	Total
All respondents	2,840	3,206	2,805	3,180	12,031
Pupils with disabilities					
Unable to hear	0	0	0	0	0
Unable to speak	4	4	3	4	15
Unable to see	2	2	5	3	12
Unable to write	11	2	8	3	24
Total	17	8	16	10	51

7.5   School location

7.5.1   Rural schools

Is there a rural/urban divide in the performance of students on literacy and numeracy tests? In order to analyse this, we supplement the data from CS3 with information from the ASC on whether the school was 'urban' or 'rural', and use this to disaggregate average pupil IRT test scores (Table 23).

Table 23:   Pupil test scores by rural/urban location

School-level mean (%)	Urban	Rural	Significant diff.	Absolute diff (urban/rural)
L2	469.7	432.2	Urban	37.5
L4	477.9	424.0	Urban	53.9
N2	474.3	424.8	Urban	49.5
N4	492.2	436.5	Urban	55.7

It can be seen from the table that there are vast differences between the performance of students in rural schools and urban schools, with students in urban schools performing significantly better than students from rural schools. The differences are more pronounced for literacy and numeracy tests in Grade 4, as compared to Grade 2. The differences are also more pronounced for numeracy tests than for literacy tests.

Does this rural–urban divide in test scores change with the amount of ESSPIN Output 3 intervention received? In order to analyse this, we disaggregate the results by the amount of ESSPIN intervention received by schools, i.e. minimum, medium or maximum (Table 24).

Table 24: **Pupil test scores by rural/urban location and amount of ESSPIN Output 3 intervention**

Mean test score (%)	Minimum intervention (zero to one years)			Medium intervention (two to three years)			Maximum intervention (four to five years)		
	Urban	Rural	Sig. diff	Urban	Rural	Sig. diff	Urban	Rural	Sig. diff
L2	445.5	425.5	Urban	487.3	447.4	Urban	553.1	468.5	Urban
L4	451.4	418.2	Urban	493.9	433.9	Urban	528.1	463.9	Urban
N2	457.2	423.2	Urban	483.4	422.6	Urban	544.7	454.5	Urban
N4	474.1	427.6	Urban	501.3	454.3	Urban	539.3	488.2	Urban

The results indicate that the rural/urban divide continues to persist irrespective of the amount of intervention received under ESSPIN Output 3. In fact, we find that the rural/urban divide increases with an increase in the amount of intervention received for all tests, except the Grade 4 numeracy test. However, it is important to note that these results are indicative, and not causal, as there may be many other factors contributing to these results, such as the poor quality of resources in rural areas as compared to urban areas and sociocultural factors, amongst others.

7.5.2   Distance from LGA headquarters

Pupils in schools that are relatively more remote may perform poorly due to a number of reasons, including poor quality of resources, challenges in recruiting qualified teachers, difficulties in monitoring teacher and/or school committee involvement, and rigid and inflexible school curricula that does not meet the needs and/or abilities of the students, amongst others. In order to determine the impact of remoteness on pupil performance in CS3, we supplement the data on the performance of pupils in Grade 2 and Grade 4 literacy and numeracy tests with data from the ASC, which provides information on the distance of the school from the LGA headquarters. We run a linear regression model of pupil test scores on distance of the school from the LGA headquarters. The results are reported in the table below.

Table 25: **Pupil performance and distance of school from LGA headquarters**

Pupil test scores	Distance of school from LGA headquarters	Significant coefficient
L2	-0.14	
L4	-0.21	
N2	-0.00	
N4	-0.15	

We find that there is a negative correlation between distance of the school from the LGA headquarters and the performance of the pupils. However, this difference is not significant. The gender disaggregated results of distance on pupil performance are reported in the table below.

Table 26: **Pupil performance and distance of school from LGA headquarters disaggregated by gender**

Pupil test scores	Distance of the school from LGA headquarters			
	Boys	Sig. coefficient	Girls	Sig. coefficient
L2	-0.06		-0.23	
L4	-0.10		-0.29	Significant
N2	-0.19		0.13	
N4	-0.05		-0.29	Significant

We find that there is no significant impact of distance from LGA headquarters on the performance of boys. However, for girls, we find that there is a significant and negative impact of distance from schools on performance of girls in Grade 4 literacy and numeracy tests.

7.6 Age correctness for grade

Age correctness for grade is important, as a wide variation in age groups within the same grade would lead to age-specific learning challenges, amongst others. The literature is unclear on the impact of being over-age on pupil performance. There are two opposing forces that may affect over-age pupils. Firstly, it may be the case that over-age pupils have received more preparation through private pre-schooling, and hence, are likely to perform better (Pritchett and Beatty, 2014). However, it may also be the case that over-age students perform better in lower grades, but as they progress, they may face growing pressure to drop-out. Thus, the promotion and survival rate of over-age pupils maybe lower (Education Policy and Data Centre (EPDC), 2011).

In order to understand the impact of being over-age on pupil performance in Grade 2 and Grade 4 in CS3, we look at each pupil's performance in relation to whether or not he or she is over-age. In Grade 2, pupils are expected to be seven or eight years old, and hence we define over-age pupils in Grade 2 as those who are nine years old and above. In Grade 4, pupils are expected to be nine or 10 years old, and hence we define over-age pupils as those who are 11 years old and above.

7.6.1 Profile of over-age pupils in CS3

The profile of the over-age pupils by state, and overall, is summarised in the following table.

Table 27: **Proportion of over-age pupils in CS3 overall and by state**

%	Grade 2	Grade 4
Over-age	46.2	46.5
Percentage of over-age pupils in each state by grade (%)		
Enugu	28.3	38.3
Jigawa	33.1	33.5
Kaduna	51.7	48.7
Kano	49.1	51.0
Kwara	41.9	28.7
Lagos	35.6	39.2
Note: The proportions shown are based on the CS3 sample with weights applied, so should be representative of the population.		

It can be seen from the above table that 46.2% and 46.5% of pupils are over-age in Grade 2 and Grade 4, respectively. In Kano and Kaduna, almost half of the pupils are over-age in both Grade 2 and Grade 4. The other states have varying degrees of over-age pupils, with the lowest being 28.3% for Grade 2 pupils in Enugu, and the highest being 41.9% for Grade 2 pupils in Kwara. The following table disaggregates the over-age pupils by gender.

Table 28:   Proportion of over-age pupils by gender

Over-age	Grade 2	Grade 4	Total
Boys	46.2	48.5	46.2
Girls	46.2	44.3	46.5
Note: The proportions shown are based on the CS3 sample with weights applied, so should be representative of the population.			

Overall, roughly the same proportion of girls and boys are over-age, with a greater proportion of boys being over-age in Grade 4.

Table 29:   Extent of over-age amongst CS3 pupils (% of over-age pupils)

Older by	Grade 2	Grade 4
One year	44.3	31.5
Two years	31.9	39.3
Three years	11.3	14.2
Four or more years	12.5	15
Total over-age (%)	100	100

It can be seen from the above table (Table 29) that amongst the over-age pupils in Grade 2, 44.3% of the pupils are older by one year, followed by 31.9% older by two years and 11.3% older by three years. For over-age pupils in Grade 4, 39.3% of the students are older by two years, followed by 31.5% who are older by one year, and 15% who are older by four years or more.

7.6.2   Results from CS3

How do over-age pupils perform compared to the rest of the pupil sample? The following table shows pupil test scores by their 'over-age' status.

Table 30:   Mean test scores in CS3 schools by over-age status

Mean test score (%)	Not over-age	Over-age	Significant difference
L2	456.6	461.1	
L4	442.3	477.8	Over-age
N2	439.8	479.6	Over-age
N4	454.7	492.0	Over-age

We find that over-age pupils perform significantly better than their age appropriate counterparts in all tests barring the Grade 2 literacy test.

The following table disaggregates the pupils' performance by the amount of ESSPIN Output 3 intervention, i.e. minimum, medium or maximum.

Table 31: **Mean test scores in CS3 disaggregated by over-age status and ESSPIN Output 3 intervention**

Mean test score (%)	Minimum intervention (zero to one years)			Medium intervention (two to three years)			Maximum intervention (four to five years)		
	Not over-age	Over-age	Sig. diff	Not over-age	Over-age	Sig. diff	Not over-age	Over-age	Sig. diff
L2	430.5	449.5	Over-age	488.7	474.7		512.6	514.3	
L4	420.3	459.3	Over-age	466.1	502.6	Over-age	497.6	514.6	
N2	422.3	472.4	Over-age	457.4	488.1	Over-age	511.7	519.1	
N4	427.5	482.6	Over-age	484.2	501.9		511.4	539.9	Over-age

It can be seen from the above results that amongst the schools receiving minimum intervention under ESSPIN Output 3, the over-age pupils perform significantly better than their age appropriate counterparts for all the Grade 2 and Grade 4 tests. This significant difference is reduced amongst pupils in schools that receive medium intervention, with over-age pupils performing significantly better only on Grade 4 literacy and Grade 2 numeracy tests. For schools that have received maximum intervention, over-age pupils perform significantly better only on Grade 4 numeracy tests, and the difference between the over-age pupils and the age appropriate pupils is much less great, as compared to the schools receiving minimum intervention.

8   Teacher competence: Gender differences

In this section, we present teacher competence results disaggregated by gender. Overall, 44.9% of CS3 teachers were female, with wide variation between northern (Jigawa: 12.7%) and southern states (Lagos, Enugu 80%+). Female teachers performed significantly better than their male counterparts for 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 large differences in the proportions of female and male teachers reaching competence standards. Female teachers were also significantly more inclusive than male teachers, both in terms of spatial and gender inclusiveness.

8.1   Gender composition of CS3 teacher sample

Overall, less than half the teachers across the six states are female (Table 32:   Percentage of public primary school teachers who are female32). There is wide variation between the states. In the north, Jigawa had 12.7% female teachers, with Kano and Kaduna at 22.6% and 56.4% respectively. Lagos and Enugu had the highest proportions of female teachers (81.5% and 82.7 respectively).

Table 32:   Percentage of public primary school teachers who are female

	Female teachers (%)
Jigawa	12.7%
Kaduna	56.4%
Kano	22.6%
Enugu	81.5%
Kwara	60.8%
Lagos	82.7%
Total	44.9%

Note. The proportions shown are based on the CS3 sample with weights applied, so should be representative of the population. Teachers who only teach 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 competence of teachers (Box 7). Female teachers appear to be performing significantly better than their male counterparts for almost all the logframe teacher competence criteria (

Table 33:   Teacher competence in CS3 by gender groups 33) except in the use of teaching aids, where there is no significant difference between male and female teachers. Compared to male teachers, female teachers demonstrated better knowledge of the English and mathematics curricula; used more teaching aids, excluding reading, writing and copying from the blackboard; praised more than reprimanded, and assigned more group/individual assignments. The number of female teachers who passed the English and mathematics content knowledge tests is more than twice that of male teachers; and the number of female teachers who met the teacher competence logframe standard is significantly more than the number of male teachers who met the same.

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 (bottom part of

Table 33: Teacher competence in CS3 by gender groups 33). Although the gap narrows, we still find within each state that a higher proportion of female than male teachers reaches the overall competence standard. However, this difference is only significant in Enugu and Kaduna.

Box 7: Logframe standard for teacher competence
A teacher must meet three out of four of the following criteria to meet the competence standard if he/she teaches English and/or maths. Teachers of other subjects must meet two out of three criteria (excluding 1 below).

1) Knowledge of English or mathematics curriculum (based on interview)

2) Use of at least one teaching aid during lesson observation

3) Greater use of praise than reprimands during lesson observation

4) Class organisation: assigning individual or group tasks at least twice during lesson observation (or for two contiguous five-minute blocks)

In CS2, a new stricter indicator of teacher competence was introduced. This excluded reading from or writing on, or having pupils copy from, the blackboard as use of a teaching aid. CS2 also added a fifth criterion:

5) English and mathematics content knowledge: scores at least 50% in both an English literacy and a mathematics test

The CS3 overall report reports performance of teachers on these logframe standards in more detail and breaks down results by CS1, CS2 and CS3; and across CS3 schools by Output 3 intervention groups. The CS3 state reports similarly delve into teacher competence in more detail at individual state level. The present report focuses on gender differences within CS3.

Table 33:   Teacher competence in CS3 by gender groups

Teacher competence criteria (%)	Male	Female	Significant difference
Knowledge of Eng./maths curriculum	31.7	55.8	F
Use of one or more teaching aid	98.3	98.8	
Use of one or more teaching aid, excl. read/write/copy from blackboard	62.4	79.9	F
Praise more than reprimand	84.1	91.6	F
Assigns two or more ind./group task	23.8	35.5	F
English score (%)	33.9	48.7	F
Mathematics score (%)	51.8	62.8	F
Passes English and mathematics test	24.8	53.2	F
Competence score (CS1 version)	2.5	2.9	F
Met teacher competence standard (CS1)	61.2	73.3	F
Competence score (CS2 version)	2.6	3.3	F
Teacher competence standard (CS2)	23.8	33.7	F
Strict teacher competence standard (CS2 version) by gender groups and states			
Enugu	12.3	24.6	F
Jigawa	25.5	34.1	
Kaduna	25.3	41.4	F
Kano	23.6	29.9	
Kwara	11.9	15.1	
Lagos	39.7	49.3	
Note. F = significant difference in favour of females; M = significant difference in favour of males (p < .05)			

When we compare the schools on the basis of the amount of intervention received under Output 3, i.e. minimum, medium or maximum, we find that the difference in teacher competence by gender is significant only for those schools that have received minimum or medium intervention, but not for schools that have received the maximum amount of intervention, i.e. four to five years. For schools that have received the maximum amount of intervention, we find that there is negligible difference between male and female teachers on all of teacher competence criteria. However, in the case of schools that have received minimum or medium amount of intervention, we find that there are significant differences between the competence of male and female teachers, with female teachers performing significantly better than their male counterparts on most counts. It should also be noted that some schools may have received intensive intervention when the programme was rolled out, but they may not have received sustained support in the form of follow-up training, refresher courses or mentoring visits. This lack of sustained support may affect the results.

Table 34:   Teacher competence in CS3, by gender and Output 3 intervention groups

Teacher competence criteria (%)	Minimum intervention (zero to one years)			Medium intervention (two to three years)			Maximum intervention (four to five years)		
	Male	Female		Male	Female		Male	Female	
Knowledge of Eng./maths curriculum	30.7	54.7	F	25.6	58.0	F	54.0	53.9	
Use of one or more teaching aid	97.9	98.6		99.4	98.8		99.1	99.7	
Use of one or more teaching aid excl. read/write/copy from blackboard	60.4	73.2	F	61.2	84.1	F	90.1	89.1	
Praise more than reprimands	82.3	87.4		86.9	94.6	F	92.6	96.9	
Assigns two or more ind./group task	25.3	28.4		15.7	37.1	F	41.9	49.3	
English score (%)	32.5	45.4	F	34.8	53.1	F	46.5	48.4	
Mathematics score (%)	49.7	58.8	F	54.4	66.8	F	64.3	65.2	
Passes English and mathematics test	21.3	46.2	F	28.8	62.9	F	46.7	50.5	
Competence score (CS1 version)	2.5	2.7	F	2.4	2.9	F	2.9	3.0	
Met teacher competence standard (CS1)	62.2	66.7		57.6	79.8	F	74.5	76.3	
Competence score (CS2 version)	2.4	3.1	F	2.6	3.6	F	3.3	3.4	
Teacher competence standard (CS2)	23.8	32.3		24.7	38.7	F	19.6	25.5	
Note. F = significant difference in favour of females; M = significant difference in favour of males ($p < .05$)									

8.3   Spatial and gender inclusive practices

Spatial and gender inclusiveness by teachers in classroom practices is discussed in more detail in Section 4 above; here we disaggregate the performance of the CS3 teacher cohort by gender on these two aspects (Table 35: Spatial and gender inclusiveness by teachers in CS3 schools, by gender groups 36). Female teachers perform better than their male counterparts on all indicators, except for participation per minute in class, where there is no significant difference between male and female teachers. 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 56% of their male counterparts. Female teachers also involve girls and boys proportionately (within 10%) in the class, as compared to their male counterparts: on an average, 4.2 zones participate in classes with female teachers, as compared to 3.6 zones in classes with male teachers – a statistically significant difference. Female teachers also receive a significantly higher gender equity score as compared to their male counterparts.

Table 35:   Spatial and gender inclusiveness by teachers in CS3 schools, by gender groups

Indicators (%)	Male	Female	Significant diff.
Participation per minute in class	30.6	33.5	
Involves boys/girls proportionately (within 10%)	41.7	51.0	F
Gender equity score (0=completely unequal, 100=perfectly equal)	79.8	84.7	F
Involves pupils from four or more areas of the class	56.5	70.0	F
Number of zones participating in lessons	3.6	4.2	F

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

8.4   Head teacher effectiveness: Gender differences

In this section, we present head teacher effectiveness results by gender. Overall, 39% of CS3 schools were headed by female head teachers, with wide variation between northern (Jigawa: 0.95%) and southern states (Lagos 87.6%). Across all six states, female head teachers appear to be performing significantly better than their male counterparts on five criteria. Overall, the number of female head teachers who met the effectiveness standard is almost three times that of male teachers. This result is also mirrored in ESSPIN schools categorised by the amount of intervention received under Output 3, with female head teachers outperforming their male counterparts. However, more significant differences are observed in those schools that have received a minimum or medium amount of intervention, vis-à-vis the schools that received maximum intervention. The pattern of female head teachers outperforming male counterparts is, however, not mirrored uniformly across the six states.

8.4.1   Gender composition of head teachers in CS3

The gender composition (weighted) of head teachers in CS3 schools is presented in Table 36: Percentage of schools headed by female head teachers in CS3 sample 36 below. Overall, 39% of the 735 schools visited in CS3 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 0.95% schools headed by a female head teacher, with Kano and Kaduna at 7.4 and 41.4 per cent, respectively. In the south, Lagos has the highest proportion of schools headed by female head teachers, followed by Enugu and Kwara.

Table 36:   Percentage of schools headed by female head teachers in CS3 sample

	Female
Jigawa	0.95%
Kaduna	41.4%
Kano	7.4%
Enugu	65.7%
Kwara	52.4%
Lagos	87.6%

8.5   Teacher effectiveness criteria

The ESSPIN logframe defines head teacher effectiveness in terms of seven criteria (Box 8). These reflect activities by the head teacher as well as 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 8: 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) Carried out two or more lesson observations in the past two weeks   2) Held four or more professional development meetings since the start of the 2011/12 or 2013/14 school year (NB: survey took place more than nine months into the school year)   3) School has a teacher attendance book and head teacher recalls at least two actions taken to promote teacher attendance   4) Clear school opening time: more than 50% of pupils sampled agree on the school opening time and more than 50% of teachers sampled agree on the school opening time   5) More than 50% of classes are in their classroom with their teacher within 30 minutes of school opening time   6) Length of morning break is 35 minutes or less, except in Enugu when it must be 15 minutes or less   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 CS3 overall report examines head teacher effectiveness with regard to these criteria, over time and between ESSPIN and non-ESSPIN schools. The CS3 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 CS3 schools and all six states together (Table 37:   Head teacher effectiveness in CS3 schools, by gender 37), female head teachers appear to be performing significantly better than their male counterparts for two of the seven logframe head teacher effectiveness criteria. Compared to male head teachers, higher percentage of female head teachers had carried out two or more lesson observations in the past two weeks (criterion 1) and had held four or more professional development trainings (criterion 2), had a clear opening time for schools (Criterion 4), at least 50% of learners and teachers were in class on time (criterion 5), and had an appropriate morning break (criterion 6). An average female head teacher met four out of seven criteria for head teacher effectiveness, as compared to three criteria met by male head teachers – a statistically significant difference. The number of female head teachers who met the effectiveness standard

(five out of seven criteria met) is almost three times that of male teachers and this difference is significant.

As with teachers, part of the difference in the effectiveness of male and female head teachers may simply reflect differences between states: female head teachers are predominantly found in Lagos, and very few are in found in Jigawa or Kano (see Table 36:Percentage of schools headed by female head teachers in CS3 sample 36 above). However, when disaggregate the results by the average number criteria met by head teachers by state (bottom part of Table 37), we find that female head teachers perform significantly better than their male counterparts in Kano. However, it should be noted that only 7.4% of the head teachers in Kano are female, and in Jigawa there is only one female head teacher.

Table 37:   Head teacher effectiveness in CS3 schools, by gender

Head teacher effectiveness criteria	Male	Female	Significant diff.
(1) lesson observations (%)	45.2	67.4	F
(2) professional development meetings (%)	35.8	56.8	F
(3) action on teacher attendance (%)	48.8	57.4	
(4) clear opening time (%)	6.9	21.4	F
(5) in class on time (%)	62.2	93.8	F
(6) appropriate morning break (%)	74.2	83.7	F
(7) appropriate lesson length (%)	29.2	19.5	
Number of criteria fulfilled (out of seven)	3.1	4.0	F
Effective head teacher (five out of seven criteria met)	3.7	10.5	F
Average number of criteria met by head teachers by state and gender (Figures in %)			
Enugu	4.1	4.7	
Jigawa	3.3	1 female HT	M
Kaduna	2.2	3.0	
Kano	3.4	4.4	F
Kwara	3.2	3.7	
Lagos	4.1	4.3	

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 all types of schools, regardless of the amount of intervention, with female head teachers performing significantly better than their male counterparts. However, these differences are more significant for schools that received minimum or medium amount of intervention than the schools that received maximum intervention. For schools that received minimum intervention, we find that female head teachers perform significantly better than male teachers for the following criteria: number of lessons observed, actions on teacher attendance, at least 50% of teachers and learners present 30 minutes after class starts, and clear opening times. Across the board, we find that female head teachers perform better than male teachers, but the difference is only significant in some cases. 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 are based in Jigawa or Kano.

Table 38: **Head teacher effectiveness in CS3 schools, by intervention categories and gender**

Head teacher effectiveness criteria (%)	Minimum intervention (zero to one years)			Medium intervention (two to three years)			Maximum intervention (four to five years)		
	Male	Female		Male	Female		Male	Female	
(1) lesson observations (%)	41.9	65.0	F	68.1	78.1		26.1	53.2	F
(2) professional development meetings (%)	32.3	38.8		50.2	77.5	F	36.2	50.0	
(3) action on teacher attendance (%)	44.9	63.7	F	66.3	62.4		47.1	39.0	
(4) clear opening time (%)	5.9	28.3	F	10.4	21.3	F	8.5	10.5	
(5) in class on time (%)	60.4	87.8	F	61.7	98.9	F	80.3	94.9	
(6) appropriate morning break (%)	77.4	82.9		62.3	84.8	F	72.8	83.3	
(7) appropriate lesson length (%)	31.7	17.6		11.1	12.2		44.4	34.8	
Number of criteria fulfilled (/7)	2.9	3.9	F	3.3	4.3	F	3.2	3.8	
Effective head teacher (5/7 criteria met)	3.4	13.7	F	2.8	12.5	F	7.4	2.1	
Note. F = significant difference in favour of females; M = significant difference in favour of males (p < .05)									

8.6 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 CS3 head teacher cohort by gender (Table 39). Overall, we find that there are no significant differences between female head teachers and their male counterparts for any of the indicators.

Table 39: **Head teachers' actions to improve pupil attendance in CS3, by gender**

% of head teachers who took the following actions	Male	Female	Significant diff.
Keep up-to-date registers	21.4	22.0	
Involve SBMC in finding reasons for non-attendance	59.4	52.8	
Discuss with teachers, pupils or parents about reasons for non-attendance	75.8	71.0	
Implement suggested solutions for non-attendance	13.1	14.7	
Address issues such as bullying and punishment	1.1	2.6	
Improve quality of teaching and learning	4.1	5.7	
Average number of actions taken	1.9	2.0	
Note. F = significant difference in favour of females; M = significant difference in favour of males (p < .05)			

9   Conclusion

The third and final round of the Composite Survey in 2016 examined a wide range of indicators in order to assess whether schools in ESSPIN states were improving over time, and whether ESSPIN schools were performing better than non-ESSPIN schools. The indicators examined in the report include inclusive practices in schools, SBMC functionality and actions to address exclusion, participation of women and children in SBMCs, and differences in education outcomes along various dimensions of marginalisation.

The year 2013/14 witnessed a large-scale roll-out of ESSPIN interventions, linking with most (and later, all) of the schools in the six states by working with state and local government officers. This large-scale roll-out occurred very close to the second Composite Survey, i.e. CS2, and after CS2 there was a move towards deepening of interventions amongst the schools receiving ESSPIN interventions, furthering the expectation that this would boost the performance of all schools in the beneficiary states. However, the overall situation, as per the third Composite Survey, indicates mixed results. While there is a significant and positive improvement in SBMC functionality, and in the participation of women and children in SBMCs, we find that there is no significant change in school inclusiveness, and spatial and gender inclusiveness between 2014 and 2016. There is no clear trend for head teachers' actions to address attendance in the aforementioned period. The increase in the participation of women and children in SBMCs can be attributed to the increase in women and children's training under ESSPIN Output 4, whereas the increase in SBMC functionality can be attributed to an increase in SBMC activities, resources raised, and greater community engagement.

When we compare schools that received intervention under ESSPIN Output 4 to the control schools, i.e. the schools that did not receive intervention under Output 4, we find that the former group of schools perform significantly better than the latter group, i.e. the control group. Schools receiving ESSPIN Output 4 interventions are more inclusive, including spatial and gender inclusiveness, have more functional SBMCs and are more inclusive of women and children. Within ESSPIN schools, we find that schools that have started receiving the interventions relatively recently, i.e. post-CS1 schools, perform significantly better than the pre-CS1 schools. This makes the case for continued training, support and engagement with the schools. Similar trends are observed when we disaggregate pupil performance by the amount of Output 3 intervention received.

Another intended outcome of the programme was to support marginalised groups, and to reduce discrepancies between the performance of pupils who come from marginalised groups or backgrounds, and those who do not come from marginalised groups or backgrounds. This was analysed in Part B of the report. However, we find that, overall, these differences in pupil performances are significant, and continue to persist. Boys perform significantly better than girls on almost all literacy and numeracy tests, wealthier pupils perform significantly better than impoverished pupils, those speaking a majority language perform significantly better on numeracy tests than those speaking a minority language, and those who live in urban areas perform significantly better than those who live in rural areas. There were also wide differences between the states, with girls in Lagos outperforming boys, whereas the reverse was true for other states like Kano and Kwara. When we disaggregate the results by the amount of intervention received under ESSPIN Output 3, we find that schools that receive more intervention under ESSPIN Output 3 perform relatively better than those that receive less intervention. Some of these differences might be attributed socio-political contexts, and other factors, such as difficulty in recruiting teachers, increases in enrolment rates (which puts pressure on the existing system), and difficulty in recruiting teachers and maintaining infrastructure in rural areas, amongst others.

To conclude, the programme has obtained mixed results, as it has proved to have positive and significant results in some areas. like SBMC functionality and participation of women and children in SBMCs, but, on the whole, inclusiveness and differences in pupil outcomes on the basis of marginalsation persists. The programme has improved results for the targeted ESSPIN schools, but the same is not true for the non-ESSPIN schools, or the schools that have received minimum inputs from the programme.

References

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

DFID. (2013). Extension to the Education Sector Support Programme in Nigeria (ESSPIN) (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. Education Policy and Data Centre (EPDC). 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). ESSPIN Overall findings and technical report. Report number: ESSPIN 060.

ESSPIN Community Engagement & 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.

Hafiz, M. and Farik, M. (2016). Effectiveness of Teaching and Learning Mathematics Using Children's Home Language and Cultural Tools. *International Journal of Scientific & Technology Research*, 123-127.

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 (CS2) 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 Annual School Census (ASC).

National Bureau of Statistics. (2013). Nigerian Annual School Census (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.

Pinnock, H. (2016). Education Sector Support Programme in Nigeria (ESSPIN): Inclusive Education Review, 2016.

Poulsen, H. (2009). School Based Management Committees (SBMCs) 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. and 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/.

Usman, A. (2016). SBMC Resource Mobilization Validation Study Summary Report.

Annex A   Testing disabled children

A.1   Ability to hear

All of the questions in both the 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.[13] Data collectors were trained on ascertaining hearing impairment using the following instructions.

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.

A.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.

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 say his name.
- In the literacy tests this was combined with a question to check whether the child could respond in English.

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 were subsequently skipped.

A.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, they were subsequently asked only questions that could be asked orally and that required an oral response.

---

[13] For pupils with hearing impairments, provision of only written instructions was considered to be an inappropriate substitute as all questions required verbal instructions and explanation from the data collector.

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.

A.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.

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.