EDOREN
Education Data, Research and Evaluation in Nigeria

Teacher Development Programme
Inspiring teaching, transforming learning

UKaid
from the British people

Teacher Development Programme (TDP)

Impact Evaluation of Output 1: In-Service Training

Final Baseline Technical Report, Volume II: Methods, Technical Annexes and Supplementary Analyses

Sourovi De, Gunilla Pettersson, Robert Morris and Stuart Cameron

June 2016

This evaluation is being carried out by EDOREN (Education Data, Research and Evaluation in Nigeria) under its TDP evaluation and support activities (Workstream 2) on behalf of the UK Department for International Development (DFID) funded Teacher Development Programme (TDP). The workstream leader is Sourovi De (sourovi.de@opml.co.uk), and the team leader for this evaluation is Ian MacAuslan. The client reference number for the project is 8022. The contact point for the client is Esohe Eigbike at DFID Nigeria (e-eigbike@DFID.gov.uk) and Bukola Oyinloye (olubukola.oyinloye@tdpnigeria.org) at the TDP.

Acknowledgments

Sincere thanks are due to the UK Department for International Development (DFID), particularly Education Advisers Esohe Eigbike and Laura Brannelly, Results Adviser Stuart Martindale, and erstwhile Senior Education Adviser Emily Oldmeadow, for useful discussions and advice on the mixed-methods impact evaluation design, and feedback on earlier drafts of this report.

Thanks also go to the Managing Agent, Cambridge Education: particularly Bukola Oyinloye (TDP Results and Evidence Technical Lead); Nguyan Feese (TDP National Programme Manager); Mourie Nishad (Former TDP Deputy Programme Manager); Celine Thomas (TDP Commercial Programme Manager); Abdurrahman Umar (TDP Pre-service Technical Lead); Robert McCormick (Senior Evaluation Specialist); David Johnson (Instrument Development Specialist); John Shotton (Sub-divisional Director, Mott MacDonald) and Ron Tuck (Senior Education Adviser, Mott MacDonald). Special thanks are due to the TDP state in-service coordinators for their assistance during fieldwork: Habeeb Saleh and Fati Abubakar (Jigawa); Hussaina Hassan and Mark Toscanini (Katsina); and Kalli Kori, Muhammed Raji Sheriff and Chinedu Mbonu (Zamfara).

Several members of Education Data, Research and Evaluation (EDOREN) and Oxford Policy Management (OPM) contributed tirelessly at various stages of the quantitative and qualitative baseline survey design, implementation and analysis. In no specific order, they are: Stuart Cameron (Quantitative Instrument Development Lead); David Megill (Sampling Specialist); Rachel Outhred (Learning Metrics Adviser); Adrian Beavis (Instrument Development Adviser and Peer Reviewer); Sourovi De (Project Manager and Education/Quantitative Specialist); Gunilla Pettersson (Education Lead Quantitative Analysis); Anna Duncan-Grant (Project Officer); Paul Jasper (Quantitative Methods Consultant); Ian MacAuslan (EDOREN Evaluation Lead and Peer Reviewer); Alex Hurrell (Impact Evaluation Adviser); Matthew Powell (Sampling Adviser); Peter-Sam Hill (Education Consultant); Girija Bahety (Quantitative Researcher); Robert Morris (Qualitative Researcher); Allan Findlay (EMIS Specialist); Terry Allsop and Monazza Aslam (Peer Reviewers); Saltanat Rasulova (Qualitative Research Methods Adviser); Oladele Akogun (EDOREN Country Director); Florian Friedrich (EDOREN Project Manager); Ijeoma Iroaganachi (EDOREN Project Administrator); Nandi Chetden (EDOREN Programme Officer); John Ode (EDOREN Project Driver); Dan Hooton, James Offer and Ayo Sangobanwo (Spearfish Risk Managers).

The quantitative baseline survey fieldwork was implemented by the survey team at OPM Abuja. Thanks and appreciation go to: Ifeatu Nnodu (Survey Manager); Charles Umar (Fieldwork Manager); Salawu Saheed (Deputy Fieldwork Manager); Babatunde Akano and Juste Ntiema (Data Managers); Busayo and Joshua Moriyonu, and Okpe Bartholomew (Data Assistants); and Bukunola Adekolu and Ibukun Ijaopo (Survey Administrators). Special thanks are due to the six state coordinators for the quantitative baseline survey: Musa Abubakar and Tijani Garba (Jigawa); Fahad Lawal and Timothy Oyetunde (Katsina); and Daniel Ayala and Dawooda Kognet (Zamfara). Under challenging conditions, they supervised the fieldwork in 330 primary schools in 42 Local Government Authorities (LGAs) across the three states. The quantitative data collection was successfully completed thanks to the immensely hard work of the data collectors who were kindly seconded from their jobs at various state education boards for the duration of the training and fieldwork. The qualitative baseline survey fieldwork was implemented by the following seven researchers who worked tirelessly in the field: Charles Umar, Dawooda Kognet, Fahad Lawal, Nafinatu Abdullahi, Salawu Saheed, Shafa'atu Musa, and Ahmed Abdulhamid. Thanks are also due to Jamila Sali and Fahad Lawal for carrying out the unenviable task of transcribing countless hours of interviews and translating them from Hausa into English.

Education Secretaries and teacher facilitators (TFs) at various Local Government Education Authorities (LGEAs), as well as staff at the three State Universal Basic Education Boards (SUBEBs), provided invaluable field advice, hospitality and support to the survey teams, thus ensuring smooth progress and the safety of the teams. Finally, sincere gratitude is extended to the team drivers Mukhtar, Ahmed and Nasiru for conveying the researchers safely during their work.

Needless to say, EDOREN thanks and extends warm regards to the head teachers, teachers, and pupils for their time and patience while participating in this impact evaluation baseline research.

Recommended citation: De, S., Pettersson, G., Morris, R., and Cameron, S. (2016, June). Teacher Development Programme (TDP) Impact Evaluation of Output 1: In-Service Training, Final Baseline Technical Report, Volume II: Methods, Technical Annexes and Supplementary Analyses. EDOREN, Abuja.

EDOREN
Education Data, Research & Evaluation in Nigeria

No. 2, Mafemi Crescent
Utako district
Abuja, Nigeria

Tel +234 803 722 0460
Email office@edoren.org
Website www.edoren.org

Table of contents

List of boxes, figures and tables

List of abbreviations

CAPI	Computer-assisted personal interviewing
CoE	College of Education
CI	Confidence interval
CSDP	Community Social Development Programme
DAC	Development Assistance Committee (OECD)
DFID	UK Department for International Development
EDOREN	Education Data, Research and Evaluation in Nigeria
EGRA	Early Grade Reading Assessment
EMIS	Education management information systems
ESSPIN	Education Sector Support Programme in Nigeria
FGD	Focus group discussion
GEP3	Girls' Education Project, Phase 3
HND	Higher National Diploma
ICT	Information and communication technology
IDS	Institute of Development Studies
INSET	In-Service Education Training
IRT	Item response theory
JSSs	Junior Secondary Schools
KIIs	Key informant interviews
LGA	Local Government Authority
LGEA	Local Government Education Authority
M&E	Monitoring and evaluation
MDE	Minimum detectable effect
MDG	Millennium Development Goal
MNSQ	Mean square fit statistics
N	Number of observations
NAP-SL	National Assessment Programme – Scientific Literacy
NCE	Nigeria Certificate in Education
NERDC	Nigeria Educational Research and Development Council
NGN	Nigerian Naira
NREC	National Research Ethics Committee
OECD	Organisation for Economic Co-operation and Development
OPM	Oxford Policy Management
OND	Ordinary National Diploma
P	Primary
PTA	Parent–teacher association
PTR	Pupil–teacher ratio
QAO	Quality Assurance Officer
R&E	Research and Evidence
SBMC	School-Based Management Committee
SEQAS	Specialist Evaluation and Quality Assurance Services
SLM	School leadership and management
SUBEB	State Universal Basic Education Board

TDNA	Teacher Development Needs Assessment
TDP	Teacher Development Programme
TF	Teacher facilitator
TOC	Theory of change
UBEC	Universal Basic Education Council
UNICEF	United Nations Children's Fund
USAID	United States Agency for International Development

1 Introduction

The Teacher Development Programme (TDP) is a six-year (2013–19) UK Department for International Development (DFID) funded education programme seeking to improve the quality of teaching in primary and junior secondary schools (JSSs), and Colleges of Education (CoEs) in six states in northern Nigeria. TDP is being implemented by Mott MacDonald and is operating in Jigawa, Zamfara and Katsina in the programme's first phase, which began in 2014. There are plans to extend TDP to Kano, Kaduna and Niger states in late 2016. In November 2014, education consultants from the Education Data, Research and Evaluation in Nigeria (EDOREN) finalised the evaluation framework for TDP's in-service component, which will use a theory-based, mixed-methods approach to assess if TDP's in-service teacher training model has improved teacher effectiveness and the learning levels of pupils in primary education in northern Nigeria (EDOREN, 2014).

This baseline report is organised into two volumes. Volume I is intended as a standalone report, which presents an overview of the programme and evaluation, and the baseline results for the programme's treatment and control areas. It is designed to be accessible to all readers. Volume II (this report) covers the technical and methodological details underpinning this impact evaluation, and further supplementary analyses not appropriate for Volume I, and is intended for those interested in methods, detailed statistical results, and the detailed qualitative background accounts for each case study school.

1.1 Structure of this volume

This baseline report is structured as follows. Section 2 provides an overview of the mixed-methods approach adopted by this impact evaluation, and explains how quantitative and qualitative data will be combined to evaluate any impact and mechanisms of TDP. Section 3 describes the quantitative methods underlying baseline results, the sampling strategy, and the processes of data collection, cleaning and analysis. It also provides a summary of the survey instruments, a brief discussion of the representativeness of results, and possible risks to the impact evaluation. Section 4 does the same for the methodology, sampling and data collection tools of the qualitative research component.

The annexes of this report contain the terms of reference for the evaluation (Annex A); TDP intervention factsheet for in-service training component (Annex B); contextual profiles and maps of TDP's Phase 1 states (Jigawa, Katsina and Zamfara) (Annex C); final sample design (quantitative survey) and weighting procedures (Annex D); school selection guidelines for randomisation (Annex E); notes on scaling the pupil test scores and item person maps (Annex F, Annex G); and supplementary analyses and detailed statistical tables (Annex H to Annex N).

Gender-generic language has been used throughout this report to refer to teachers, pupils and parents; where necessary, for ease of reading, the female noun and pronoun have been used to refer to all genders.

2 Mixed-methods approach

The impact evaluation framework for TDP's in-service training component outlined the strong case for a mixed-methods approach to impact evaluation (EDOREN, 2014). Mixed-methods designs are particularly well suited to the evaluation of complex programmes like the TDP, which cover multiple levels of a system (working with pupils, teachers, schools, teacher trainers, training institutions, and government system-level engagement). The evaluation plan proposed using both quantitative and qualitative surveys and the evaluation will therefore draw on different sources of data. This is important for several reasons:

- Different types of data will shed light on different types of important phenomena; the use of only quantitative data, for instance, would mean the evaluation would focus in a partial way on outputs/outcomes/impacts that can be measured quantitatively, and would be limited in regard to the types of explanations that can be drawn around these data. While the ultimate aim of a 'good' impact evaluation is to identify cause and effect relationships, qualitative data can help us to understand the mechanisms driving impact or cause and effect relationships.
- Different data sources can help to triangulate findings.
- Different data sources can inform each other, both through an iterative process, where the quantitative surveys are designed using results from qualitative research, and vice versa, depending on the sequence, and where the results of the analysis of one data source inform, and are combined with, the analysis of the other.
- Qualitative methodologies can be more flexible to ongoing changes to the programme and therefore can help to answer questions that may not be covered by the quantitative survey. For instance, if many teachers in schools in the quantitative sample move after receiving training, this may weaken the ability of a quantitative survey to determine impact, and qualitative data will then play a stronger role in the assessment.

The three main considerations in designing the mixed-methods approach for this impact evaluation therefore included: integration of methodologies for better measurement; sequencing information for better analysis; and merging findings for better action (Carvalho and White, 1997; Garbarino and Holland, 2009). Specifically, the TDP evaluation follows the following practical steps proposed by Garbarino and Holland (2009).

Integrates methodologies for better measurement

- uses indicators from the baseline quantitative surveys to help select a qualitative investigation sample;
- uses baseline quantitative survey results to highlight priority issues and to generate new hypotheses to cover in qualitative research;
- uses initial qualitative analysis (the reviews of the literature and stakeholder consultations conducted in developing the evaluation framework) to identify knowledge gaps to be filled by the quantitative surveys;
- uses initial qualitative analysis to prioritise issues that are important to stakeholders that should be covered by the quantitative surveys; and
- uses initial qualitative analysis to construct indicators.

Sequences information for better analysis

- generates working hypotheses from an initial qualitative study, to test using quantitative methods;
- uses a qualitative study in a sub-sample of quantitative areas to compare findings;
- uses a qualitative study to assess heterogeneity behind quantitative averages from the surveys;
- uses a qualitative study to explain relationships emerging from the quantitative surveys; and
- uses a qualitative study to triangulate quantitative findings.

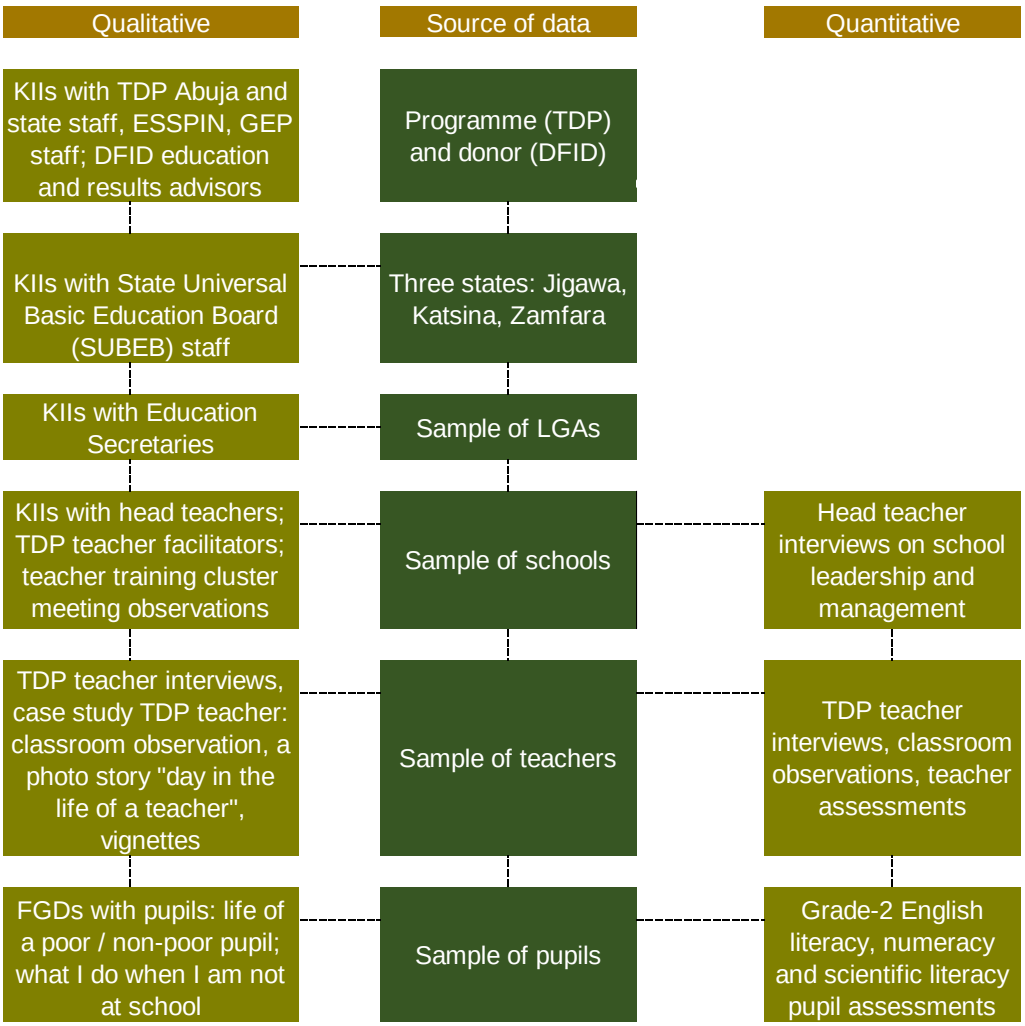
Merges findings at analysis stage into a single, mixed-methods report

- produces an integrated mixed-methods report drawing on both quantitative and qualitative data.

2.1 Combining qualitative and quantitative data

Figure 1 illustrates how the quantitative and qualitative data sources will be combined to yield an understanding of (any) TDP impact and potential mechanisms through which the programme may have influenced the outputs and outcomes of interest. It shows that while the quantitative survey focuses on collecting data at the school level, covering head teachers, teachers and pupils in a representative sample of programme impact evaluation areas, the qualitative research collects data in a small purposive sample of states, local governments, and schools and from TDP in-service teacher trainers at the regional level. The quantitative and qualitative components are designed to complement each other to allow for deeper insights into any impact and mechanisms of TDP.

Figure 1 Quantitative and qualitative data collection



Source: EDOREN impact evaluation team

The timing of the qualitative fieldwork was adjusted so that the first part would take place in June 2014 (key informant interviews (KIIs) with programme staff at TDP Abuja and state offices; with Education Sector Support in Nigeria (ESSPIN); with Girls' Education Programme (GEP); and with DFID education and results advisers), and the second part in June 2015 (state, Local Government Authority (LGA), school, teacher and pupil interviews, and focus group discussions (FGDs)) after the completion of the quantitative fieldwork (October–November 2014). This was done to allow the qualitative fieldwork to explore selected initial findings from the quantitative research with head teachers, teachers and pupils. As such, part of the qualitative research programme was left open, so as to be flexible to the emerging findings from the quantitative survey. This impact evaluation collects primary data as it is not possible to measure programme impact on pupil learning outcomes and teacher professional capacity using secondary sources (such as the Annual

School Census or the Nigeria Education Data Survey), as these are not available or they fail to capture the core areas the programme is seeking to impact.[1]

The quantitative data collection and initial analysis took place before the qualitative design was finalised, in order that the findings from the quantitative baseline could feed into the qualitative research design to enable it to explore issues uncovered by the preliminary quantitative data analysis and factors that contribute to observed outcomes at school level. Specifically, the qualitative research guides for the school and community qualitative research were finalised after the completion of the quantitative baseline survey, with some questions being based on the preliminary quantitative findings.

As a result of this staggered sequence and the timing of the quantitative and qualitative surveys, the evaluation was also able to conduct a 'light touch' process evaluation undertaken as part of the baseline qualitative study, which took place roughly six months after commencement of implementation of in-service training activities. This gave the authors an opportunity to study various implementation processes and practical dynamics, and in turn to provide initial impressions of how implementation of the in-service output is progressing. Having said this, by their very nature it is atypical of baseline surveys to observe and comment on the implementation of an intervention. Thus, the core focus of this baseline survey report is to establish baseline levels of teacher effectiveness and pupil learning before the start of TDP's in-service teacher training activities, especially comparability between the programme's treatment and control groups, which will be used to evaluate programme impact at the follow-up rounds of data collection in 2018.

[1] The baseline mainly uses primary data for the analysis. Secondary data were used for sampling (Annex D: Sampling strategy and sample size in the present Volume II). Some existing literature and data were used to understand the intervention context – these are referenced in the bibliography.

3 Quantitative impact evaluation design

This section starts by describing the methodology for the quantitative approach, including identification of the control group, the sampling strategy and the sample size. It then discusses the development of the survey instruments and discusses each instrument. Finally, it outlines risks to, and possible limitations of, the quantitative approach.

3.1 The counterfactual, causal attribution and randomisation

This impact evaluation is designed to answer the question as to whether TDP's in-service teacher training programme has resulted in improvements in both teacher effectiveness and in pupils' learning over time. One of the key challenges in designing a robust impact evaluation, in order to establish a causal relationship, is to define a valid counterfactual – measuring what would have happened in the absence of the intervention. In this case, this means identifying control groups of schools, teachers and pupils that do not participate in the TDP in-service training but that with respect to relevant background characteristics are, on average, similar to the treatment groups of schools, teachers and pupils that will participate in the TDP. For example, teachers in the treatment and control schools should have similar qualifications, teaching experience, age and so on, such that the only difference across the two groups is ideally with respect to whether a teacher participates in the TDP or not.

The method this impact evaluation adopts to create a counterfactual is to *randomly* assign *clusters* of schools to either the treatment or the control group.[2] An important point to note is that the random assignment is not done at the level of individual schools (or teachers or pupils), but for clusters of schools. This is because, by design, the TDP combines schools into groups of 12 schools in each LGA, based on their geographical proximity to each other.[3] Three teachers and the head teacher are, then, based on pre-defined criteria, selected within each treatment school (in the cluster) to participate in the TDP,[4] while no teacher in a control school (in the cluster) receives TDP's in-service training.

On the basis of the assumption that the random assignment of treatment worked as intended, any potential difference in teacher effectiveness and pupil learning in the treatment and control group over time can be attributed to the TDP.[5] To assess the validity of the counterfactual the baseline results given in Volume I are presented overall, as well as by treatment and control group, together with tests for significant differences in group means.

[2] Thus, in this impact evaluation and baseline survey, the unit of assignment to treatment/control is the school, while the unit of analysis is the teacher or pupil.

[3] This is to make the TDP more effective: for example, by facilitating cluster meetings and peer learning among teachers and head teachers who are receiving the in-service training.

[4] TDP thus works with a limited number of teachers per school (three teachers and the head teacher), and, based on the size of the school, the proportion of teachers per school reached by the programme varies considerably. For instance, the (unweighted) mean number of teachers employed in a school was 12 as per this baseline survey, meaning that on average 33% of total teacher strength is enrolled in the programme in treatment schools. However, this varies widely as schools in the 10th percentile (in terms of teacher numbers) have only one teacher, while those in the 90th percentile have 57 teachers in a school. Thus, in some smaller schools the programme is reaching 100% of teachers, while in larger schools it may be reaching 10% or less.

[5] It can be argued that the simplicity of experiments offers considerable advantages in providing convincing results for researchers and policy-makers. It can also be argued that random assignment is the fairest and most transparent way of choosing the recipients of a new programme. There are, however, fierce debates around these views. See, for example, Duflo and Kremer (2004), who are proponents of this approach, and Deaton (2009), who argues against (or at least remains cautious about) using randomised control trials to evaluate development effectiveness.

3.1.1 An overview of the randomisation process

For the selection of the treatment and control groups the EDOREN evaluation team recommended to DFID Nigeria and TDP that within each of the 14 local (LGAs) in each of the three TDP Phase 1 states the TDP should select two clusters of 12 schools[6] each. To keep the school selection process straightforward, the EDOREN evaluation team provided SUBEBs with guidelines that laid out a set of school characteristics to consider when selecting the clusters (see Annex E). These guidelines aimed to achieve the selection of two school clusters in each LGA that were 'as similar to each other as possible', but without being too prescriptive about the characteristics to balance them on.[7]

To prevent bias in the selection of teachers within the selected treatment and control schools, the identification of teachers who would participate in the TDP had to occur before the treatment and control school clusters were selected. Thus, within each LGA, schools in both treatment and control groups were required to select four teachers each (before knowing if the school would participate in the TDP or not), who would potentially benefit from the TDP. The programme's criteria for selecting teachers are shown in Box 1. In every school (treatment and control) the programme always selected the head teacher (whether they teach or not), as well as three teachers.

Box 1 Criteria for selection of teachers to participate in TDP in-service training

- Classroom teaching at early grade level (Grades 1–3); and
- Classroom teaching in any of the three subjects: English, maths, and science.

The initial design aimed to assign school clusters within each LGA to treatment or control status through a state-level public randomisation process witnessed by members of the relevant SUBEBs, LGEA, head teachers and TDP staff, overseen by the EDOREN evaluation team. This would constitute the study population of 12 treatment and 12 control schools in each of the 14 LGAs in the three TDP Phase 1 states. However, due to security concerns, which prevented the EDOREN evaluation team travelling to the TDP states, as well as due to time constraints, the public randomisation process could not be carried out. Instead, the EDOREN evaluation team, after receiving lists of school clusters and teachers from the TDP, assigned clusters in each LGA to either the treatment or the control group using a simple random-number generator.

3.1.2 State-specific randomisation processes

In Jigawa, since nearly all schools had received ESSPIN interventions by September 2014, the TDP in-service working group selected two clusters of 12 schools each in 14 LGAs by trying to establish a balance in the level of exposure to ESSPIN across the two clusters. Each set of schools was clustered to allow for regular cluster meetings and further peer learning among the trained teachers. Generally, two ESSPIN clusters (each of which have five to seven schools in them) were

[6] To be specific, the impact evaluation design adopted here is one of constrained randomisation. A fully randomised evaluation of TDP would have required all public primary schools in the states in which TDP operates to be allocated into clusters. Within each state a large number of clusters would have been randomly selected to participate in the programme (treatment group), with a similar number randomly selected as the control group. Crucially, the selection of the four TDP teachers within each school would have had to occur after the treatment and control groups had been selected, but before they had themselves been informed of their treatment status.

[7] Balance between treatment and control school clusters could not be checked *ex ante* because of a lack of (reliable) data on schools and teacher characteristics.

combined to form a single cluster of 12 schools for TDP. As mentioned above, the EDOREN evaluation team, after receiving lists of school clusters and teachers from the TDP coordinators in Jigawa, assigned clusters in each LGA to either the treatment or the control group.

In Katsina and Zamfara, where there are no pre-existing clusters of schools (these are not ESSPIN states), the in-service working group selected two sets of 12 schools each in 14 LGAs, taking into consideration their distance from each other for ease of classification into clusters. Both these states are also selected for the GEP Phase 3 (GEP3), a programme which was undergoing re-design at the time schools were being selected for this baseline survey. As mentioned above, the EDOREN evaluation team, after receiving lists of school clusters and teachers from the TDP coordinators in Katsina and Zamfara, assigned clusters in each LGA to either the treatment or the control group. At the time of drafting this report, the EDOREN evaluation team was assured that both GEP3 and TDP were clear that each of their interventions would not be provided in schools covered by the other programme. This is a situation that will be monitored continuously.

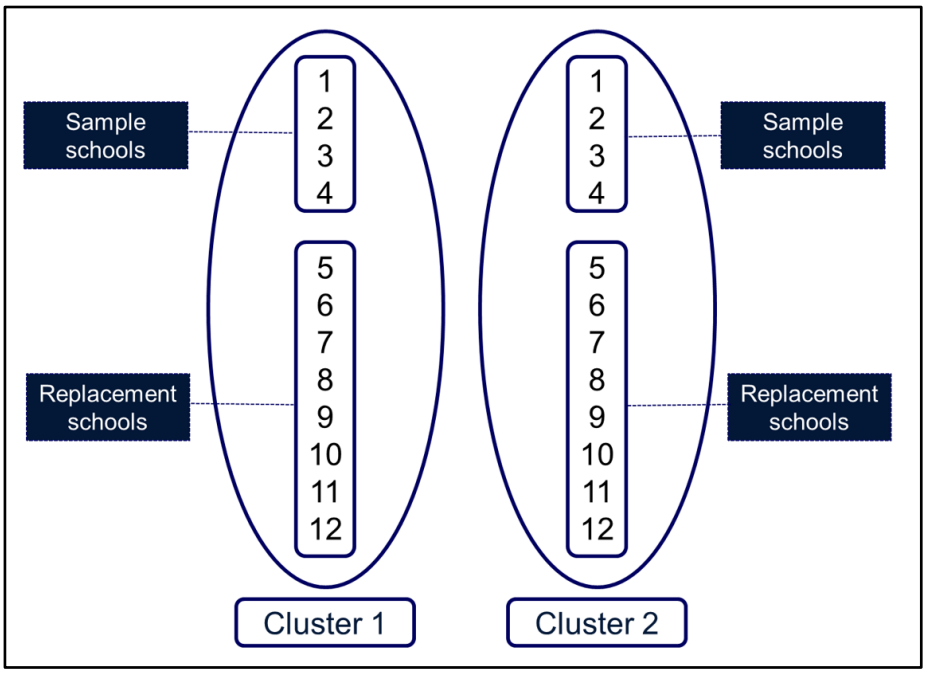
3.2 Sampling strategy and sample

3.2.1 Sampling strategy

The overall sampling strategy was shaped by practical programme considerations and resource constraints. Within these parameters, the design is intended to maximise the statistical power of the impact indicator difference-in-difference measures (and to reduce the minimum detectable effects (MDE) so as to make the evaluation as sensitive as possible to detecting small changes). A technical note detailing the sampling strategy can be found in Annex D below.

The TDP will operate in 14 clusters per state. As set out above, there are also an additional 14 control clusters in which schools went through the TDP teacher selection process but where the programme will not operate. Clusters consist of 12 schools and from every cluster four schools were randomly sampled for the quantitative baseline survey, as shown in Figure 2. This yields a total of 112 schools (56 treatment and 56 control) per state, and 336 schools (168 treatment and 168 control) in total for the three TDP Phase 1 states.

Figure 2 Organisation of treatment and control clusters within each LGA



Since each cluster consists of 12 schools, and only four of these were to be sampled, the remaining eight schools were available as replacement schools in case the sampled schools could not be surveyed. However, the process of school replacement was closely monitored and controlled, to avoid introduction of school selection bias.

Sampled schools were allowed to be replaced only if any of the following situations arose:

- the school was not eligible for TDP intervention (e.g. it was an integrated Quranic, Tsangaya education school, or it was a special school for children with a disability);
- the school was closed for the duration of the survey team's stay in the LGA;
- there were security concerns about visiting the school; or
- the school did not have any TDP / control teachers who teach English, maths or science to Grade 3 pupils.

At each school, the (one) head teacher and (three) selected teachers were interviewed. Each teacher and head teacher who teach were also observed while they taught a class. Following the completion of the school survey, all teachers and head teachers (irrespective of whether they teach or not), were administered a teacher development needs assessment (TDNA) at an examination centre.

In order to assess pupil learning levels for this baseline survey, eight pupils among all those who started Grade 3 in September 2014 and who were being taught English, maths or science by at least one TDP/control teacher, were randomly selected for the combined English, maths and scientific literacy learning assessment.[8] The pupils were drawn from a sampling frame consisting of all eligible Grade 3 pupils present in school on the day of the survey by data collectors, using a random-number generator programmed into their computer-assisted personal interviewing (CAPI) software.

[8] The rationale for administering a combined subject assessment to the sampled pupils is both pragmatic and technical but has the risk that some of the supposedly 'treatment' pupils may not be taught by a TDP teacher in all three subjects.

The quantitative component of this impact evaluation follows a longitudinal design.[9] The same cohort of teachers, head teachers, and pupils who were surveyed at the baseline will be surveyed again at the endline in June 2018. This will allow measurement of the impact of the TDP on teacher effectiveness, and on pupils' learning between Grade 3 and Grade 6. While the pupils who had recently started Grade 3 at the time of the baseline survey in October 2014 were administered a Grade 2-level learning assessment, they will be administered a Grade 6-level assessment at the endline in June 2018 (when they will be coming to the end of Grade 6) that will include a limited number of the same Grade 2-level items to maintain direct comparability with the baseline.[10]

3.2.2 Intended sample sizes

The intended total sample sizes for all three states are: 336 head teacher interviews; classroom observations of head teachers who teach (up to 336); 1,008 teacher interviews and classroom observations; 1,344 TDNA administered to both teachers and to head teachers; and 2,688 Grade 3 pupils tested on Grade 2 learning assessments (Table 1).

Table 1 Summary of intended sample sizes

Instruments/evaluation clusters	Per school	Treatment (per state)	Control (per state)	Total (per state)	Total (three states)
Evaluation clusters	n.a.	14	14	28	84
Total sampled schools (given that four are sampled per cluster)	n.a.	56	56	112	336
Head teacher interviews	1	56	56	112	336
Head teachers *who teach*: classroom observations	Up to 1	Up to 56	Up to 56	Up to 112	Up to 336
Teacher interviews	3	168	168	336	1,008
Teacher classroom observations	3	168	168	336	1,008
TDNA (teachers and all head teachers)	4	224	224	448	1,344
Pupil learning assessment	8	448	448	896	2,688

3.2.3 Sample coverage

The intended sample for the quantitative baseline survey is 336 schools. During the fieldwork 330 schools, i.e. 98% of the intended sample, were surveyed. Five sample schools could not be surveyed or replaced due to a lack of eligible schools in the treatment/control cluster where selected TDP/control teachers were teaching English, maths or science to Grade 3 pupils. In one LGA in Katsina, one school could not be surveyed or replaced due to security concerns (Table 2).

[9] Further follow-up data will come from surveys by EDOREN using the same instruments, and also the same fieldwork management teams to the greatest extent possible. The same 330 schools surveyed at baseline will be surveyed again in the endline (the pupils and teachers are also panelled). The likelihood of some of the schools closing down over the four-year period is small. Attrition of pupils and teachers is likely and conservative estimates of attrition were assumed in sample size calculations when the quantitative survey sample was being determined.

[10] The academic school year in Nigerian public primary schools begins around September/October, and lasts for three terms, ending in June/July in the following year.

Table 2 Intended and actual sample sizes and response rates

Instruments	Expected per school	Total intended	Total actual	Response rate
Total sampled schools		336	330	98%
Head teacher interviews	1 per school	336	330	98%
Teacher interviews	3 per school	1,008	908	90%
Classroom observations: teachers and head teachers who teach	3 teachers and head teacher if teaches	1,226	1,077	88%
TDNAs (teachers and all head teachers)	4 (3 teachers and 1 head teachers)	1,344	1,158	86%
Grade 2 pupil learning assessment (administered to Grade 3 pupils)	8 Grade 3 pupils per school	2,688	2,575	96%

Within the sample schools it was not always possible to administer all instruments to the intended number of respondents, for a range of reasons. In some cases this was because the school was very small and lacked sufficient numbers of eligible pupils and/or teachers. Where head teachers did not teach,[11] classroom observations were not conducted for these head teachers. In a small number of cases sampled pupils and teachers left the school before being tested/interviewed/observed – for example, due to illness. Finally, the head teachers and selected teachers from seven of the sample schools did not show up at the examination centres for the TDNA. Reasons for less than full response are expected to equally affect both treatment and control clusters and thus this is unlikely to affect the randomisation design. However, the sample as a whole might be subject to selectivity bias if the schools, teachers and pupils who were ultimately included in the sample are systematically different from the rest of the population of TDP treatment and control schools, teachers and pupils.

3.2.4 Revisits

As mentioned above, for every school in a TDP/control cluster, names of the head teacher and three selected teachers were provided to the evaluation team, and thus no teacher sampling procedure was conducted in the schools during the survey. However, these lists were not up-to-date in many cases, which meant that a listed teacher could not always be found in the school on the day of the survey. There are many possible reasons for this, including short-term absence (e.g. due to illness, training), long term absence (e.g. due serious illness, study leave), transfer, death and teachers who were unidentified. Most cases of short-term absences were later covered during school revisits by the survey teams.

Teachers who were transferred, or could not be identified, could not be included in the survey. In order to preserve the teacher sample size, after consultation with the programme it was decided that data collectors would ask the head teacher to name a replacement teacher, as per the selection criteria given in Box 1, and these new teachers (from treatment schools) would then be

[11] To give a sense of this in terms of percentages, in 112 of the 330 schools visited (i.e. 34% of the school sample) the head teacher was not a regular teacher and was hence not subject to classroom observation.

enrolled for the in-service training. EDOREN has shared with TDP the database of replaced teachers, for their enrolment into the programme.

3.3 Survey instruments

The quantitative baseline survey administered five different instruments in the treatment and control schools using CAPI. The TDNA was separately administered on paper, to mimic real-life marking of pupil tests and preparation of worksheets, to sampled teachers and all head teachers at examination centres after the completion of the quantitative baseline survey.

The head teacher and teacher interviews were conducted in Hausa, the predominant language spoken in northern Nigeria. The pupil numeracy and scientific literacy assessments were also administered in Hausa, including a number of the English literacy items – except those where comprehension of the question in English was essential to the competency being tested by the item.[12] The TDNA was a written test and was administered in English. Table 3 lists the quantitative instruments and respondents, and provides brief descriptions of the instruments' contents.

Table 3 Overview of instruments and respondents

Instrument description	Respondent(s) per school
Head teacher interview and school record checks	
Head teacher gender, age, years of experience, academic qualifications, training undertaken	Head teacher (one per school)
Frequency/type of interaction with and supervision of teachers	
Number of pupils registered and teachers employed, pupil–teacher ratio (PTR), school infrastructure and resources, School-Based Management Committees (SBMCs)	
Teacher attendance from school records	
Portrait photograph of head teacher for panel identification at follow-up rounds of survey	
Teacher motivation questions (administered to those head teachers who teach)	
Teacher interview	
Teacher gender, age, years of experience, academic qualifications, training undertaken	Three sampled teachers
Frequency/type of in-service training received	
Interaction with and supervision by head teacher	
Self-reported absenteeism	
Teacher motivation	
Portrait photograph of teacher for panel identification at follow-up rounds of survey	
TDNA in English, maths and science and technology	
Assessment of subject knowledge and ability to measure and analyse pupils' academic progress	Same three sampled teachers and head teachers

[12] For instance, checking whether the pupil could answer a simple question was done by asking 'What is your name?' in English. If the skill being tested was writing, however, then instructions (e.g. 'Please write the first letter of your name') could be communicated in Hausa.

Instrument description	Respondent(s) per school
	(irrespective of whether they teach or not)
Classroom observation	
Key teacher behaviours in the classroom, including teacher talk, teacher language, teacher actions recorded at the end of every three minutes	Same three sampled teachers, and head teachers who teach
Pupil activities in classroom recorded at the end of every three minutes	
Instances of praise and reprimand by the teacher; use of teaching aids; etc.	
Pupil learning assessment English literacy, numeracy and science	
English literacy: early literacy, reading with comprehension, writing, other	Eight randomly sampled Grade 3 pupils
Numeracy: pre-numeracy and Grade 1 and 2-level numeracy questions	
Scientific literacy: pre-science and Grade 2-level questions	
Pupil gender, age, language and household assets	
Portrait photograph of pupil for panel identification at follow-up rounds of survey	

3.3.1 Measuring pupil learning levels

The logframe indicator corresponding to improved pupil learning levels is *percentage change in knowledge and comprehension of three core subjects of primary school pupils in Grades 2 and 5.* To collect data on baseline levels of pupil learning a combined English literacy, numeracy and scientific literacy assessment was administered to eight randomly sampled pupils from Grade 3 in each treatment and control school. Every sampled pupil was assigned a data collector who tested the pupil within the school compound but away from her class, teacher and head teacher. The combined English literacy, numeracy and scientific literacy test on average took 40–45 minutes to complete and pupils were given colouring pencils, juice and biscuits at the end of the assessment to thank them for their time and effort.[13]

At the time of the baseline survey in October/November 2014 the pupils were about two to four weeks into the new academic year in Grade 3 and therefore the assessment included items from Grade 2 or below. The administered items can be mapped to the national curriculum set by the Nigerian Educational Research and Development Council (NERDC).[14] The English literacy and numeracy items in the pupil assessment were adapted from the Grade 2 learning assessments used for ESSPIN's[15] biennial composite school survey. These assessments have been used to collect pupil learning data for ESSPIN's baseline in 2010 and follow-up surveys in 2012 and 2014, respectively. The scientific literacy items in the test were jointly developed by EDOREN's education

[13] No other gifts or remuneration were given to pupils, teachers, head teachers or the schools.

[14] NERDC, among other core activities, develops the curriculum and instructional materials for public primary, junior secondary and senior secondary levels in Nigeria (see http://nerdc.ng/).

[15] The DFID/UK Aid-funded Education Sector Support Programme in Nigeria (ESSPIN) aims to improve learning outcomes for children in primary schools in six states of Nigeria (Jigawa, Kano, Kaduna, Enugu, Kwara, Lagos), with a range of activities at the state, national, local and school level. At the school level, it provides and supports the use of structured teaching materials, trains head teachers in academic leadership and school improvement planning, and facilitates community participation through school-based management committees.

consultants and TDP's instrument developers, based on the Grade 2 science and technology curriculum prescribed by NERDC and the Universal Basic Education Commission (UBEC).

Apart from Grade 2-level items, the combined assessment also includes items to assess pupils' foundational skills in English literacy, numeracy and scientific literacy. These include: (1) new items developed for this baseline survey to assess pupils' pre-literacy, pre-numeracy and foundational scientific literacy[16] skills; and (2) items based on the Grade 1 curriculum for these subjects from the ESSPIN assessments.

Box 2 Balancing between 'floors' and 'ceilings' in a pupil assessment

The motivation for including foundational skills items in the TDP pupil assessment was to examine pupils' learning profiles using items with varying levels of difficulty. The information from these items would also help shed light on pupils' ability to engage with the Grade 2 curriculum, particularly for those pupils who may have acquired these foundational skills but who are struggling to make the transition to develop skills deemed appropriate for a Grade 2 pupil, as well as for those Grade 2 pupils who may not have acquired even foundational skills.

These foundational items are also included to help reduce **floor effects** (Beavis and Outhred 2014), which arise when most pupils taking an assessment find the items too difficult such that a large proportion of pupils score near the bottom of the scale, also referred to as 'clumping near zero'. While adding foundational-level items increases the likelihood of detecting the extent to which pupils have acquired some of the more rudimentary literacy, numeracy and scientific literacy skills, including these items introduces the commensurate risk of subjecting the test to **ceiling effects**, whereby a large number of pupils find the test to be relatively 'easy' overall and achieve high scores (leading to clumping near the top of the scale). As discussed below, it was difficult to know, before the test was implemented and the data analysed, if the test would succeed in avoiding floor effects or if it would lead to consequent ceiling effects. This was especially pertinent for the pre-literacy and pre-numeracy items, as well as the entire scientific literacy test, which was developed especially for this survey and was being administered for the first time to pupils in this baseline phase of the evaluation. As discussed below, the test appears to have largely avoided both these effects, i.e. there is no significant clumping at either zero (floor effects) or 100% (ceiling effects).

Table 4 maps the items by subject, learning domain and grade level, and provides a snapshot of the various tasks pupils were asked to perform during the assessment. Note that the literacy test was an English literacy test, so the terms English literacy and literacy are used interchangeably in this report. Furthermore, the term 'foundational skill items' refers to pre-literacy, pre-numeracy, foundational scientific literacy and Grade 1-level items.

[16] The scientific literacy test did not contain any items that could reasonably be classified as reflecting Grade 1 level. Thus, items either test foundational science skills or knowledge of the Grade 2 curriculum.

Table 4 Mapping of subjects, learning domains and grade levels in the TDP pupil test

Learning domains	Foundational level	Grade 2 level
English literacy		
Early reading and literacy (includes both pre-literacy and Grade 1-level items)	Answering simple oral questions using complete sentences (e.g. 'What is your name?'); alphabet knowledge; print concepts related to writing; foundational oral vocabulary; knowledge of letters or sounds; knowledge of initial sounds of words; reading familiar words; knowledge of initial sounds of words; answering simple questions in complete sentences.	
Skills for reading with comprehension		Listening with comprehension; reading a passage aloud; spelling familiar words; writing answers to oral questions
Writing		Spelling familiar words; writing plurals; writing answers to oral questions
Other literacy skills		Pupil's use of continuous present tense; using plurals
Numeracy		
Number concepts	Number recognition, counting to 10, sequencing/ordering numbers	Counting above 100; filling up missing numbers in a sequence; fractions
Addition and subtraction	Addition of one- or two-digit numbers; subtraction of numbers less than 20; buying items and receiving change for amounts equal to, or less than, Nigerian Naira (NGN) 10.	Addition and subtraction of two-/three-digit numbers; buying items and receiving change for amounts up to NGN 500
Other numeracy skills	Measuring lengths using non-standard methods (palm, steps); comparing capacity; recognition of 2-D shapes	Multiplication of single-digit numbers; measuring length using standard methods (ruler); measuring capacity; reading time on the hour/half-hour
Scientific literacy		
Scientific literacy skills	Comparing length, weight and tastes; matching colours	Understanding processes (dyeing a cloth); movement of objects through air/water, work/effort/machines

Source: TDP in-service baseline survey (October 2014), pupil learning assessment. Note: The classification of literacy and numeracy items by learning domains closely follows ESSPIN's mapping, though some items could be testing multiple learning domains. See ESSPIN 2013, Annex C, for details.

At endline (June 2018) the same pupils, who will then be at the end of Grade 6, will be given a Grade 6-level English, maths and scientific literacy assessment to measure any impact of TDP's in-service training output. The TDP logframe expects a three percentage point improvement[17] in mean test scores for pupils in the TDP treatment schools between Grades 3 (2014) and 6 (2018).

The rest of this section focuses on results from the baseline survey on the overall performance of pupils in treatment and control schools in English literacy, numeracy and scientific literacy; by grade levels, learning domains, and pupil background. The analysis uses estimates of pupil achievement based on Rasch modelling, which makes it possible to validly compare learning assessments across grades and over the years.[18] As such, a scaled score is the mathematical transformation of a pupils' raw scores in order to report her score on a continuum consistently over the years and across different versions of the assessment (Bangladesh Directorate of Primary Education 2013; Outhred 2015). Rasch analysis also allows for test difficulty and pupil ability to be reported independently on the same scale. In addition, a scaled score of, say, 500 will mean the same at endline in 2018 as it did at baseline in 2014.[19]

Raw scores are frequently used to calculate pupil learning performance by analysing the percentage of test items that the pupil answered correctly, or the percentage of pupils answering one or more items correctly. Although this is arguably the easiest way to grade tests, from a statistical point of view raw scores suffer from several drawbacks (Outhred 2015). First, when calculating a percentage score each assessment item is given an equal weight, which implicitly assumes that every item is equally indicative of the level of knowledge or skills that the pupil possesses. In reality, this is likely not to be true. The use of raw scores also limits the extent to which assessment results can be compared over time and across locations. Finally, raw scores confound the difficulty level of the assessment with the ability or knowledge level of the pupils. It is not possible to separate out these two components, which jointly determine the percentage of questions a student answers correctly. Bearing these relative merits of scaled versus raw scores in mind, scaled scores have been used for the main analysis while raw scores have also been computed. These are presented in Annex M for the reader's reference.

For each subject, pupils are sorted into three performance levels, each described by a set of competencies expected at Grade 1 or Grade 2 levels, or below. This: gives insights into whether pupils are performing at or below the curriculum level expected at Grade 2; helps identify relative strengths and weaknesses in terms of specific skills; and indicates if specific groups of pupils (boys and girls, poorer and richer) perform differently. An advantage of using scaled scores it that it makes is possible to assess changes along the pupil learning distribution. That is, whether the programme moves pupils primarily from Level 0 (below Grade 1-level skills) to Level 1 (at Grade 1-level but below Grade 2-level skills), primarily from Level 1 to Level 2 (Grade 2-level skills), or both. For example, the programme might shift a substantial group of pupils from Level 0 to Level 1, but leave the proportion of pupils at Level 2 relatively unchanged, or vice versa. This is important to understand as certain groups of pupils may be more likely to benefit from the TDP.

[17] The impact evaluation is currently designed to measure a change in average raw test scores. As discussed in EDOREN 2014, this is being interpreted as a 3 percentage point increase in the mean test scores of pupils taking the test in 2014 and 2018. As mentioned above, discussions are ongoing to finalise the timing of further survey rounds.

[18] Rasch modelling was used to generate estimates of pupil 'ability' in English literacy, numeracy and science on an interval scale that is directly linked to grade-level competencies based on the NERDC/UBEC curriculum. The Rasch model is probabilistic and is a special case of an item response theory (IRT) model.

[19] Use of Rasch modelling for education assessments is not without its critics and the debate over its technical rigour and efficacy is ongoing; see, for instance: Goldstein 2015; Panayides, Robinson, and Tymms 2010.

3.3.1.1 Scientific literacy

Compared to English literacy and numeracy assessments, scientific literacy is not as commonly covered in learning assessments, particularly in developing countries. In discussion with education and learning metrics specialists the EDOREN evaluation team decided to use the classification of scientific literacy items and competencies as described in the National Assessment Programme – Scientific Literacy (NAP-SL), which is administered to a sample of Australian pupils every three years (Donovan *et al*. 2008). NAP-SL lists the following 'strands' or domains as capturing the process of acquisition of scientific literacy skills acquisition. The strands are:

- Strand A: formulating or identifying investigable questions and hypotheses, planning investigations and collecting evidence.
- Strand B: interpreting evidence and drawing conclusions from their own or others' data, critiquing the trustworthiness of evidence and claims made by others, and communicating findings.
- **Strand C: using science understandings for describing and explaining natural phenomena, and for interpreting reports about phenomena.**

Post hoc matching of items to strand descriptions show that all the TDP scientific literacy items can be classified under Strand C. Roughly half the science items were taken from the Grade 2 science textbook, as set by NERDC-UBEC, while the rest target foundation-level scientific literacy skills. Topics covered in the assessment include size, weight, colour and taste discrimination, the motion of objects in water and air, work and effort, and everyday chemical processes relevant to the northern Nigerian context (dyeing cloth). All the items in the scientific literacy assessment could be answered orally in either Hausa or English.

3.3.2 Measuring teachers' subject knowledge

3.3.2.1 Overall approach

The approach to the measurement of teachers' subject knowledge in the quantitative baseline was closely founded on the teacher assessment framework, TDNA instruments, benchmark of expected teacher professional working knowledge and levels of achievement framework developed previously by a reference group of national educators and international experts for two education programmes currently being funded by DFID in Nigeria, namely ESSPIN and GEP (Johnson, D. 2008; Johnson and Gabrscek 2008; Johnson and Hsieh 2014).[20]

Drawing on Johnson and Hsieh (2014), the TDP TDNA covers subject knowledge in English, maths and science, and also tests teachers' ability to assess and monitor pupils' academic progress. The TDNA has four parts, as described below and in Table 5.

[20] As noted in Johnson and Hsieh (2014), the ethics of testing teachers remains a contested issue, and has ethical implications: 'Testing teachers (like testing children) can be a contentious issue, not least because of the perceived validity of the tests. Until now the usual practice … has been to give teachers the same tests (normally mathematics) as those devised for primary aged children … likely to undermine both the validity of the tests as well as teacher status and professionalism … many teachers feel insulted … do not understand the nature of the task in relation to what they have to do as professionals. In other words, these kinds of tests have poor face validity and are fraught with ethical problems.'

Table 5 Mapping of TDNA topics, exercises and questions

Subject	Topics covered
English	Reading for factual information; interpreting words, phrases or sentences; writing a model letter in English
Maths	Addition, subtraction, multiplication, division, fractions, decimals, time, measurement, unit conversions, reading graphs, square roots, exponents
Science	The lifecycle of insects, changes in weather, functioning of the human digestive system, measurement, the process of evaporation, forms and sources of energy
Assessing and monitoring pupils' academic progress	Analysis of pupil test scores, including making simple graphs to monitor performance, short assessment, pointing out the strengths and weaknesses in each letter. Commenting on the following: purpose, organisation, grammar, spelling and punctuation
Source: TDP baseline survey (October/November 2014), TDNA instrument.	

The TDNA was made up of a series of tasks that approximated closely what teachers did or were expected to do in their everyday classroom practices, all of which emulated the ordinary pedagogical practices of teachers, as described below.

To assess **knowledge of English for teaching primary grades** teachers were asked to write a model letter in English, and to read two newspaper articles in English, from which they were to extract factual information, and to demonstrate comprehension and an ability to summarise information[21] (Table 5).

To assess teachers' **knowledge of maths** the TDNA asked teachers to mark a maths test completed by a Grade 4 pupil (age 10). This test covered addition, subtraction, division, decimals, computing time, measurement, unit conversions, fractions and reading of graphs. If a pupil's answer was correct, a teacher was expected to mark the answer correct, and if the pupil's answer was incorrect the teacher was expected to mark it incorrect and provide the correct answer.

To assess **subject knowledge in science** the third part of the TDNA asked teachers to prepare worksheets for Grade 4 pupils on a number of grade-appropriate science topics, such as the lifecycle of insects, the functioning of the human digestive system, and forms and sources of energy. These questions were based on the Grade 4 curriculum, as designed by NERDC.

The TDNA also assessed **teachers' ability to assess and monitor pupils' academic progress**, which was done through two exercises. First, the TDNA asked teachers to summarise information on pupils' test performance and to present pupils' progress in a graph. Second, it asked teachers to read, comment upon and identify relative pupil strengths and weaknesses based on three 'fake' letters written by Grade 4 pupils.

Thus, rather than ask the teachers to 'take the test' that may have been administered to pupils, they were asked to mark test papers. As Johnson and Hsieh (2014) argue that the end result in such a form of testing, in terms of assessing the teacher's subject knowledge, remains the same, in

[21] Using the newspaper articles, teachers then had to prepare model answers for an English test for Grade 6 (12 year old) pupils.

that a teacher will need to calculate the answers to the mathematics questions or to know the correct form of language in order to 'mark' the test papers.

The TDNA was administered in English and therefore there is a possibility that it not only assessed subject knowledge and ability to comprehend pupil progress but, to some extent, teachers' English language skills too. The maths section would be relatively unaffected by this, as items were numerical, with minimal instructions (in English). The English exercises aimed to assess subject knowledge of English and needed to be administered in English. For the science and pupil assessment exercises, the language could potentially be an issue in that these items might seem like tests of language competency in English, rather than tests of knowledge of science or ability to assess pupils. Nonetheless, the official medium of instruction in Nigeria from Grade 4 onwards is English (Federal Republic of Nigeria 2004), and thus teachers would be expected to understand the basic assessment instructions in the TDNA and to be able to answer questions in English.

3.3.2.2 Validity of the TDNA instrument

Given that there are no standard norms by which to assess teachers' levels of achievement, the original test developers worked with a reference group made up of senior educators in Lagos State to set the levels of the achievement framework (Johnson and Hsieh 2014). The reference group participated in the development of the test papers and tasks, and were thus familiar with the nature of what was being asked of teachers. The basic guiding questions for those who themselves were involved in the education system in Nigeria were 'reasonableness' and 'importance'. What percentage of the questions and tasks presented to teachers could a teacher reasonably be expected to master, given the nature of the material that they were being asked to work with? How important was proficiency in the material to the way in which children were taught and what they learned?

Bringing the best available principles of criterion-referenced assessment to bear on this question, a number of levels of achievement, complete with descriptors of achievement, were developed. Then, using the guiding questions of reasonableness and importance given above, an optimum professional working knowledge threshold was agreed upon.

3.3.2.3 TDNA levels of achievement

To examine differences in teachers' subject knowledge, four achievement levels were defined for the GEP TDNAs (Johnson and Hsieh 2014). These were used for the TDP TDNA too (Table 6).

Achievement Level 1 captures teachers who have 'sufficient professional knowledge' (TDNA score 75%–100%): that is, they meet the benchmark minimum knowledge threshold and are considered to be effective in the classroom. At **achievement Level 2** teachers have 'near-sufficient professional knowledge' (TDNA score 50%–74.9%) and would benefit from some in-school support and in-service training to make them effective in the classroom. **Achievement Level 3** includes teachers who have 'emerging professional knowledge' (TDNA score 25%–49.9%) and who, to be effective, would require a combination of school-based in-service training and more fundamental professional development. Finally, teachers at **achievement Level 4** (TDNA score 0%–24.9%) have limited professional knowledge and would need substantial and sustained training and support in order to become effective in the classroom.

Table 6 TDNA levels of achievement, descriptors and score ranges

Achievement level	Descriptor	TDNA score (%)
Level 1	Sufficient professional knowledge	75<=score<=100
Level 2	Near-sufficient professional knowledge	50<=score<75
Level 3	Emerging professional knowledge	25<=score<50
Level 4	Limited professional knowledge	0<=score<25
Source: Johnson and Hsieh 2014		

3.3.3 Classroom observation descriptors and scoring scheme

The types of teacher talk and action and pupil activity recorded during the classroom observations and whether defined as a positive action and item descriptors are shown in the tables below.

Table 7 Classroom observation: Teacher talk/action and pupil activity descriptors

Code	Talk / action / activity	Practice descriptor
Teacher talk		
a	Instructs / presents / dictates to the whole class	Teacher talks to the whole class but does not question or give feedback. S/he might give instructions or dictate a text for pupils to write, or 'present' some text directly from the textbook or the blackboard. For example: • Instructs — 'Today we will learn about shapes. Open your notebooks and copy these drawings of shapes from the blackboard'. Note that classroom management instructions such as 'sit down', 'stand up and raise your hands', etc. will not count as instructions here. • Presents — Reading directly from a textbook or blackboard without any additional own explanation. This could include reading out a story, poem or a passage. • Dictates — 'Open your notebooks and start writing as I say, "There are seven, s-e-v-e-n, days in a week"'; or 'Today I will give you a spelling test. Write in your notebooks. First word is umbrella…"um-bre-lla". Second word is machine…"ma-sheen"'.
b	Leads whole class, chants	The teacher asks pupils to repeat what s/he has said, thus leading to chanting by the whole class. This could include repeating what the teacher has said; a poem; chorus song; or the whole class reading out aloud from a text. Chants are usually preceded by the teacher saying 'Say/repeat after me'. For example, the teacher says: 'Repeat after me…Today is Monday…', and the pupils repeat together, 'Today is Monday'.
c	Asks a closed question or gives a closed response to the whole class	Closed question / response — The teacher asks a closed question, which has only one right answer or just a brief answer, usually for the pupils to remember facts; or answers pupils' questions in a way that closes the conversation, even if the pupils' question was an open one. For example: • Teacher: 'Children, what is the capital of Nigeria?'; pupils: 'Abuja', or 'the capital of Nigeria is Abuja'. Here there is only one right answer. • Teacher: 'Children, tell me…do you enjoy coming to school?'; pupils: 'Yes teacher, we/I do', or 'No teacher, I/we don't'. Though both answers are correct, they are brief and this closes the conversation. • Pupil: 'Teacher, how can I grow up to become a doctor?' Teacher: 'Study hard'.
d	Asks an open question or gives an open response to the whole class	The teacher asks a question that has many possible answers so that pupils imagine or analyse; or answers pupils' questions in a way that invites further discussion or thought, even if the pupils' questions were close-ended. For example: • Teacher: 'Children, why do you like coming to school?'; pupil 1: 'Because I like meeting my friends'; pupil 2: '…because I like to read books'; pupil 3: "…because I want to study and be a doctor when I grow up.' • Pupil: 'Teacher, how many states does Nigeria have?'. Teacher: 'Ok, that's a good

Code	Talk / action / activity	Practice descriptor
		question…let us try to answer it together. Each of you will name a state and I will write it on the blackboard and we will then count. Aminu, tell me the name of a state'. Aminu: 'Kaduna'. Teacher: 'Good. Aisha, tell me the name of another state…'
e	Assists individuals or groups / joins group discussion	The teacher helps groups or individuals, or joins pupils' discussions. This may typically involve the teacher moving around individual pupils or groups, stopping to check on them and assist them.
f	Explains how something works / how to do a task[22]	Teacher explains how something works or how to do a certain task, often (but not necessarily) using a teaching aid. This is different from giving instructions in the sense that it does not involve telling pupils what to do, but rather how to do it, and it typically involves breaking down single activities or concepts into smaller, easier sub-activities. For example: • A teacher may say: 'Do this addition sum: 3+2=?'. This is an instruction, while an explanation would mean s/he says: 'Here are three apples, and here are two more apples. When we put them all together, they add up to 1-2-3-4-5, five apples. So 3+2 is equal to 5'. • Teaching drawings or explaining scientific processes (with or without the use of models or equipment) will usually be categorised here. For instance, the teacher may say, 'Today you will all draw a duck. Look at my drawing on the board and try it yourself. First draw the beak, then the head. Give it an eye and then draw the neck…' • An explanation is also different from presenting, as defined in (1) above. While presenting here is defined as reading directly from the textbook or blackboard, explanation would mean the teacher adds content to the text from her/his own to make the text/concept simpler for the pupils.
Teacher action		
a	Writes on / reads from blackboard	The teacher writes on or reads from the blackboard: this could be writing mathematical exercises from the textbook on to the blackboard, homework assignments, or simply reading what has been written on the blackboard.
b	Demonstrates / displays work using the blackboard	Teacher uses the blackboard to explain a concept or problem, show how to solve a mathematics problem, illustrate a grammar or spelling point; or holds up a pupil's exercise book to explain something.
c	Moves around among students	Teacher moves away from front of the class and may look at pupils' work or join group discussion. Generally, this is not accompanied by any of (a), (b), (d) or (e).
d	Uses materials (printed/improvised that teacher has made)	Teacher uses printed materials, or observes while pupils use them under her guidance. Improvised materials include something that the teacher has made.
e	Uses textbook to explain something / reads from it	Teacher explains something from the textbook; explains a task in the textbook; or reads from the textbook.
Pupil activity		
a	Group or pair discussion / presentation	Pupils are organised into groups or pairs and discuss a topic, or report back on the results of a group discussion or group work.
b	Group or pair work to complete a task	Pupils are organised into groups to complete some task.
c	Respond to open question	One or several pupils respond to a question that has many possible answers and invites discussion. See discussion above on open questions under teacher talk for examples.
d	Respond to closed question	One or several pupils respond to a question that only has one right answer or that can be answered with a brief response. See discussion above on open questions under teacher talk for examples.
e	Individual work	Pupils work on their own tasks individually, exercising independent thought in the process, e.g. completing exercises set by the teacher.

[22] This action was added after the TDP classroom observation design note was written.

Table 8 Classroom observation: Teacher talk/action and pupil activity scoring scheme

Code	Activity	Defined as positive action?	Score	Activity	Defined as positive action?	Score	Activity	Defined as positive action?	Score
	Teacher talk			**Teacher action**			**Pupil activity**		
a	Instructs / presents / dictates to the whole class	Never	0	Writes on / reads from blackboard	Never	0	Group or pair discussion / presentation	Always	0.5
b	Leads whole class, chants	Never	0	Demonstrates / displays work using the blackboard	Always	1	Group or pair work to complete a task	Always	0.5
c	Asks a closed question or gives a closed response to the whole class	Never	0	Moves around among students	Always	0.5	Respond to open question	Always	1
d	Asks an open question or gives an open response to the whole class	Always	1	Uses materials (printed/improvised that teacher has made)	Always	0.5	Respond to closed question	Never	0
e	Assists individuals or groups / joins group discussion	Always	1	Uses textbook to explain something / reads from it	Always	0.5	Individual work	Never	0
f	Explains how something works / how to do a task	Always	1						
g	None of the above	Never	0	None of the above	Never	0	None of the above	Never	0

Note: The behaviours in each of the three categories teacher talk, teacher action and pupil activity are mututally exclusive and exhaustive. Each category (teacher talk, teacher action, pupil activity) will be scored separately. The minimum score for each interval is 0 and the maximum score is 3. For the analysis the overall score for each teacher is rescaled to obtain a total score between 0 and 1.

3.3.4 Measuring teacher attendance

The baseline survey collected teacher absenteeism data from two sources: teacher attendance records that include all teachers at the school, as well as interviews with the three TDP (or control) teachers at each school who self-report their absenteeism. All except two of the sampled schools had teacher attendance records. The TDP logframe teacher absenteeism indicator, average daily teacher absenteeism over the previous five days, is based on school records (while interview data are used for triangulation).[23,24]

Due to the timing of the baseline survey – two weeks into the second term – the maximum possible recall period was 10 working days. Nonetheless, as the first week of term may not be representative of a regular school week, a five-day recall period was chosen. One issue to note is that the absenteeism data do not distinguish between authorised absences (for instance, where a teacher has permission to be absent to attend training) and unauthorised absences (for example, where a teacher is absent without permission, to engage in, say, other income-generating activities). Nevertheless, regardless of whether teacher absence is authorised or unauthorised the implication is the same: a loss of instructional time, with adverse effects on pupil learning.[25] Finally, it should also be noted that the data collectors' visit to the school on the day of the survey was not unannounced – while individuals teachers may not have known of the survey team's visit, the head teacher or assistant head teacher were generally informed a day in advance.

Box 3 Comparing school records and self-reported data on teacher absenteeism

The baseline survey also asked teachers directly if they had been absent (see above) over the previous five working days, and, if so, for how many days. Based on the self-reported data, average daily teacher absenteeism was 8%, compared to 14% based on school records. This discrepancy (6 percentage points) in teacher absenteeism between the two sources is likely to be due to under-reporting of absenteeism by teachers, as being absent is considered a negative behaviour. The fact that the school records and the self-reported data capture different groups of teachers (all teachers at the school versus only TDP/control teachers, respectively) may also account for part of the discrepancy.

3.3.5 General note on instrument validity and reliability

If sufficient time had been available to carry out a number of pre-tests, and if large enough test samples were available, along with sufficient time afterwards to analyse the data collected from each sample, it would have been possible to discuss instrument validity and reliability in a more statistically robust manner. However, this was not possible due to time and budgetary constraints. Researchers therefore ensured all instruments reflected certain fundamental (even if basic) properties of validity and reliability, as described below.

Face validity: To ensure that the quantitative instruments largely measured what they intended to measure, and that the questions were phrased appropriately and response options were reasonable, the questionnaires were reviewed by TDP, and by several EDOREN consultants, and were piloted and revised several times before final data collection. It should also be noted that

[23] One reason for using school records data rather than self-reported data is that teachers might hesitate to report absenteeism because being absent (unauthorised) is considered a negative behaviour.
[24] The indicator is constructed as the total number of teachers absent over the previous five school days divided by the total number of teachers employed over the previous five school days, multiplied by 100.
[25] Unless there are replacement teachers in the case of authorised teacher absences.

many, if not all, items (especially for the pupil assessment) come from ESSPIN composite surveys, which have gone through several iterations through large-scale surveys over the years, to address this validity criterion.

Content validity was addressed by a similar mechanism. For instance, EDOREN consultants added some pre-literacy questions to the pupil tests to obtain richer data on pupils' learning at the lower end of the scale, and for this EDOREN researchers reviewed the existing literature to understand what exact learning domains (e.g. alphabet recognition, phonic knowledge) could be tested. They then developed a set of pre-literacy items which catered to the said domains.

Finally, researchers ensured **inter-rater reliability**, i.e. consistency among interviewers, through rigorous and uniform training for all data collectors and their supervisors, and through the provision of manuals with detailed guidance on how to ask each question. This was particularly important for the pupil tests and lesson observations.

3.4 Data collectors' training and fieldwork

To ensure consistency in administering various interviews, tests and observations, rigorous and uniform training for all enumerators and their supervisors was conducted, using detailed enumerator manuals as reference material (particularly for the pupil tests and lesson observations). Fieldwork managers and supervisors were recruited on the basis of their previous experience as data collectors and/or supervisors of large quantitative surveys in northern Nigeria, in most cases with school-based components. A number of them had previous experience of working on the ESSPIN composite surveys. Data collectors were staff seconded from the SUBEBs of the three TDP Phase 1 states for five weeks of data collection, selected on the basis of a written quiz which tested survey skills and experience and IT skills. Data collectors were especially trained to collect high quality data while protecting the identities and interests of vulnerable groups (e.g. disabled children).

The enumerator training and data collection were carried out from September through to November 2014, while data cleaning and analysis were conducted from December through to March 2015. The training, data collection, data cleaning and analysis were carried out as shown in Table 9.

Table 9 Timeline for data collection, cleaning and analysis

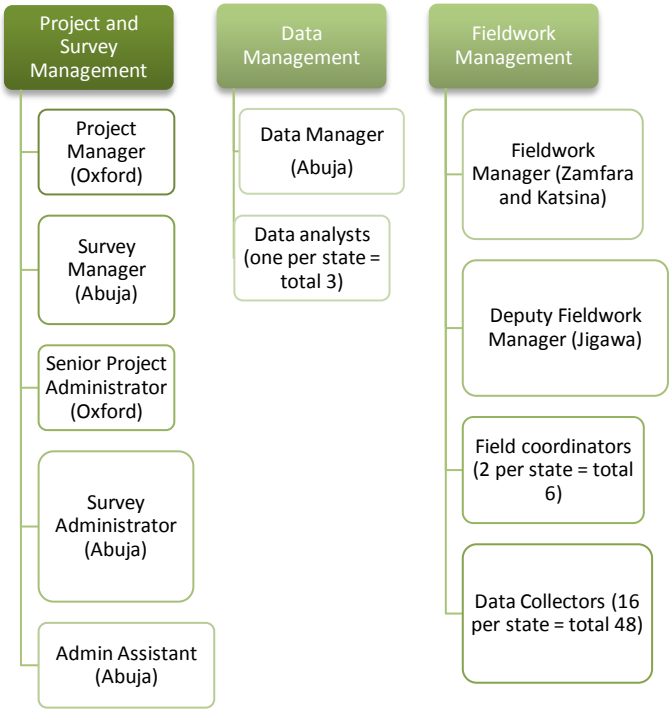
Activities	Dates	Location
Training for fieldwork manager, deputy fieldwork manager, and state coordinators	18–28 September 2014	Abuja
Field pilot 1	30–31 September 2014	Kaduna
Training for data collectors	10–18 October 2014	Abuja
Field pilot 2	16 October 2014	Kaduna
Field pilot 3	20–23 October 2014	Kaduna
Data collection	27 October–13 November 2014	Schools in Jigawa, Katsina and Zamfara
School revisits (as needed)	14–21 November 2014	Schools in Jigawa, Katsina and Zamfara
TDNA	29 November and 6 December 2014	Examination centres in Jigawa, Katsina and Zamfara
Data cleaning, analysis and report writing	5 January–31 March 2015	Oxford
Submission of quantitative baseline survey report and state reports	April and May 2015	Oxford
Presentation of quantitative baseline survey findings to DFID, TDP and TDP annual review team	October 2015	Abuja
Submission of quantitative findings integrated into a mixed-method report to TDP evaluation steering committee	November 2015	Abuja
TDP evaluation steering committee meeting to discuss feedback and future of the evaluation	January 2016	Abuja
Receipt of detailed feedback from DFID and TDP	February–March 2016	Abuja
Finalisation and submission of mixed-methods report to DFID for SEQAS review	April 2016	Oxford/Abuja

Source: EDOREN evaluation team

Data collection was not carried out by a third party but rather by OPM/EDOREN, who were involved in the TDP impact evaluation from the very beginning, including the design of the evaluation framework, design of the instruments, piloting, training, fieldwork supervision, data cleaning, analysis and reporting. Field supervision arrangements were instituted to ensure high

quality data collection. The project, data and fieldwork management structures are shown in Figure 3 below.

Figure 3 The project, data and fieldwork management structures



3.5 Data cleaning and analysis

Quantitative data were collected in the field using CAPI. Field coordinators were trained to download data daily from enumerators' CAPI devices and to transmit them via the internet to survey data managers in Abuja. Data managers, in turn, conducted checks on the data daily for errors like incomplete questionnaires, incorrect school/teacher/pupil IDs and duplication in unique IDs. These were reported back to the field coordinators for rectification the next day while they were still in the same LGA, and before they moved to the next LGA on their fieldwork plan. Field coordinators were also trained to complete a survey monitoring form (one per school) to provide more qualitative information, such as information about what time the team reached the school, whether any issues were experienced in accessing the school, whether and why teachers were replaced, reasons for less than expected number of respondents in the school, the need for revisiting the school (say, if a sampled teacher is ill and away), and so on. Data were transmitted to the lead quantitative researcher on a weekly basis for running further data checks while the team was still in the field – these checks included checking for inconsistent values not picked up by CAPI, which the field teams were then requested to clarify with respondents on the phone or by revisits during the final/contingency week of fieldwork. In summary, a large part of the data cleaning task was completed and issues rectified on a continuous basis while the teams were still in the field. At the end of fieldwork, further data cleaning was conducted by data managers (e.g. in relation to duplication of sampling units, incorrect unique identifiers, etc.) and passed on to the lead researcher. In terms of the quality of the data received, these were in a fairly workable condition, partly due to CAPI checks being built into the questionnaires and partly due to the daily checks by data managers and rectification in the field.

Prior to commencement of quantitative data analysis, an analysis plan was prepared by the lead researchers and confirmed by TDP. This contained information about what indicators would be constructed and reported, what disaggregations were required and feasible, how to present the structure of the report, etc. This plan was used as the basis for analysis and reporting. The raw data received from data managers were taken through a final round of cleaning if required. Intermediate variables were prepared whereby indicators were created. Finally, weights were applied to the intermediate data to produce overall and disaggregated means, and other summary statistics. All group means were tested for significant differences – statistically significant differences between groups are marked with asterisks in tables, as follows: *significant at the 10% level; **significant at the 5% level; ***significant at the 1% level.

All the quantitative analysis was carried out by an education/quantitative specialist using Stata version 11, while the advanced analysis for scaled pupil test and teacher motivation scores were carried out by an external IRT specialist.

Analysis coding using Stata took into account the specific sampling design features of the data: the *Svyset* command was used to declare the sample design features of data, specifying weights using the school as the primary sampling unit *(psu)*; and *strata* were specified for each school as a unique combination of state/LGA/treatment status. In cases where a strata may contain only one school (e.g. if others schools were disqualified and not replaced) the *singleunit(scaled)* command was used. Since the four teachers selected to the programme in some cases constituted a large proportion of total teacher strength in the school (e.g. in small schools), the analysis specified *fpc.* FPC or finite population correction is the proportion of primary sampling units sampled within each stratum (used only for sampling without replacement).[26] For pupils who were sampled within schools (which were sampled themselves) second-stage sampling (ssu) was appropriately incorporated in the analysis coding.

It is important to note that these reports have used "cluster" in an operational/programmatic sense (borrowing from TDP or ESSPIN) which refers to a collection of schools grouped together by the implementers for logistical purposes. However these "clusters" are not the same as the term "clusters" normally referred to in traditional sampling methodology. Rather than being sampling units drawn from a larger population in order to represent that population, ***the 84 "clusters" and the schools in them are the whole population that the survey data represents***.[27] Please see Annex D.5 (Final sample design and weighting procedures) in Volume 2 of the Evaluation report:

> *"The stratification of the sampling frame for the TDP baseline survey is by individual treatment or control cluster, since an independent sample of schools was selected from each cluster in the frame. In this case these are not 'clusters', in classic sampling terminology – actually, each primary sampling unit (school) is a cluster of teachers and pupils. Each 'cluster' of schools is considered to be a separate stratum, and the study population consists of all the TDP and control clusters in the three states. Within each cluster stratum, the individual schools are sampling clusters of teachers and pupils."*

[26] Due to clustering in the selection procedure, individuals are not selected independently. This results in correlation within clusters that can inflate variances of estimates compared to those obtained from a simple random sample of the same size, hence the need for finite population correction.

[27] This particular feature of the external validity or generalisability of baseline survey results is detailed in both volumes of the report: volume 1 (section 2.4.4) and volume II (section 3.6).

However, there is undoubtedly clustering of pupils and teachers in each sampled school.[28] However, we have attempted to allow for these clustering effects at pupil- and teacher-levels in the analysis, as explained above. Thus, by using the *svy* command and further specifying the primary sampling unit (or psu) as the school, the analyst sets the survey settings in STATA to allow for homogeneity (and therefore ensuring precision in standard error calculation) resulting from clustering of pupils and teachers who are selected from any given school. With just over 50 treatment and control schools in each state there should be enough school-clusters for this to work at the aggregate and state-level.[29]

Apart from overall mean, essential summary statistics, wherever relevant, have been disaggregated by gender, treatment/control, household assets quintiles ('richest'/'poorest'), and state. Supplementary analyses have also been provided (e.g. disaggregating of teacher characteristics for those with the highest TDNA scores). These are provided in Section 3 onwards in Volume I, and in Annex I onwards of the present report. Three individual state-level baseline reports (quantitative results only) discuss state-level results in more detail, hence the current set of reports (i.e. Volume I and Volume II) only provide summaries. Further cuts of data sub-sets, though interesting, were not always possible to calculate robustly, due to small sizes[30] of sub-samples (e.g. pupil test scores by gender in the top/lowest income quintile) or due to the lack of reliable secondary data to disaggregate the data (e.g. rural/urban status of schools). All summary statistics (especially those relating to the programme logframe) are disaggregated by treatment/control status in Section 3 onwards in Volume I. Furthermore, supplementary statistics tables are provided in Annex S of this report. To conserve space and enhance readability, where treatment/control differences are not significant these have not been dwelt on in detail but rather are summarised upfront, with a note such as: 'There were no significant differences between treatment and control schools/teacher/pupils.'

3.6 Generalisability of baseline results

The impact evaluation survey results are representative of the TDP treatment and control populations. However, these populations are not in themselves representative of the three Phase 1 TDP states more broadly. This is because the TDP clusters were purposively chosen by SUBEBs in each LGA, rather than being randomly sampled from a comprehensive list of potential TDP clusters (into which all schools can be allocated).

Furthermore, in both treatment and control schools, if treatment (or control) teachers did not teach Grade 3 pupils the entire school was replaced during fieldwork. This is because the evaluation's focus is on the impact of the TDP in-service teacher training on pupil learning levels, as measured by changes for the Grade 3 cohort (EDOREN, 2014). Replacing schools in this way

[28] Teachers are also clustered in schools but not sampled. Instead head teachers simply nominated the teachers who would be selected for the pilot if their school was in a treatment cluster (or as "control teachers" if their school was a control school in the pilot).

[29] Within each stratum, however, there are at most four clusters so the svy command would definitely not work. However, both for reasons of data protection/anonymity and sample size constraints, there are no plans of publishing stratum level estimates although advanced techniques (e.g. wild-p bootstrapping approach) could perhaps be used as a robustness check, if not for publishing stratum-level results at the endline.

[30] It was decided, as a rule-of-thumb, that disaggregations with less than 50 sampling units would not be carried out due to the fact that tests of significance become unreliable as sample sizes decrease.

could introduce a bias,[31] albeit one that is balanced across treatment and control schools, meaning that the impact results would not be representative of TDP impact overall but only of schools where the selected teachers taught Grade 3. However, it was decided that this potential risk is outweighed by the need to maintain the treatment and control pupil and teacher cohort sample sizes.

Due to these risks of bias in the data, it will be important to understand (through analysis and by consulting stakeholders) how the TDP schools differ from other schools in Jigawa, Katsina and Zamfara in general.[32] This will help us to make judgements about the extent to which the impact evaluation results are a good indicator of the likely impact of TDP if it were scaled up in these states. In turn, qualitative assessments will be made as to whether the impact evaluation results can be reliably generalised to the Phase 2 TDP states, as well as to other states in Nigeria more generally.

3.7 Achieving balance across treatment and control groups

The baseline results given in Volume I show that the randomisation approach was successful, i.e. on average, the treatment and control groups are balanced in their characteristics. This was established by checking for significant differences in means using standard t-testing, as well as tests for joint orthogonality for key variables following (McKenzie 2015). Thus, the control group may serve as a valid counterfactual for the treatment schools (and teachers and pupils) for measuring programme impact on the key outcomes of interest (changes in pupil learning levels and in teacher effectiveness).

It is worth noting that there are two indicators where differences between treatment and control groups were found to be statistically significant: pupils per class (treatment 39.6 vs control 44.3) and average % of teachers' absenteeism from school (treatment 12.9% vs control 14.8%). Could these differences potentially lead to the treatment group showing better results (defined as a 5 percentage points increase in the proportion of pupils performing at grade level in English literacy, numeracy and scientific literacy, as per the programme logframe) independently of the TDP intervention?
Although it is certainly conceivable in theory that the extra teacher time per pupil suggested by these two statistics might have such an effect, it is unlikely to fructify in reality, for two reasons.

First, in a learning environment characterised by whole class teaching, small changes in class size are unlikely to have practical significance to translate directly into differences in pupil's learning experience. Even statistically, in both cases (class sizes and teacher school absenteeism) the effect sizes are very small (Cohen's D effect size here is 0.03 to 0.05) and thus differences are likely to be negligible in practice, as noted in Volume I of the report.

Second, as the qualitative research findings clearly shows (section 4.3.4, Volume I) the particular teaching and learning environments captured by these baseline results are marked by substantial *classroom absenteeism* whereby teachers might be present within the school compound but missing in action from classrooms. Therefore, presence or absence of teachers *in school*, is not the

[31] For instance, it could be hypothesised that it is mainly in larger schools with many teachers that TDP (control) teachers were selected, such that none teaches Grade 3 pupils.
[32] The process of comparing TDP schools to other schools in the three states is, admittedly, not straightforward. For Jigawa, ESSPIN composite survey results (which are representative of the entire state) could be used. For Katsina and Zamfara, the comparative analysis is likely to rely on the Annual School Census data for state-level comparators.

most important factor determining learning but rather it is the presence or absence of teachers *in the classroom*. This further downplays the importance of the differences in school absenteeism between treatment and control groups. Indeed both quantitative and qualitative methods can be used at the endline to determine whether class size and teacher absenteeism differences were significant drivers of any changes in learning outcomes.

In summary, the minor differences in two indicators do not violate the overall validity of the control groups as a counterfactual for the treatment groups, especially when the effect sizes of the differences in question are not large enough to bear any potential practical significance to the learning and teaching environment within which this study is based.

Nevertheless, although the baseline sample is balanced, the issues outlined below pose risks to the impact evaluation. The main issues and associated risks are summarised below. In addition, we discuss what can be done to preserve the impact evaluation design and to enable it to measure programme impact on pupil learning levels and teacher effectiveness over time.

3.8 Risks to the quantitative impact evaluation

Some key features of this quantitative survey's design, namely randomised assignment to treatment and longitudinal survey/panelling of teachers and pupils, pose risks to its implementation over time. These mainly relate to contamination of control clusters by TDP or other programmes, and ensuring that as many of the teachers and pupils sampled for the baseline as possible remain in the study for the endline surveys. Changes to the programme design and timeline after the implementation of the baseline survey, as well as the timing of the follow-up rounds of surveys, also constitute risks to the impact evaluation in terms of what it will be able to measure at the endline. These risks and their implications for the impact evaluation, and what would need to be done to address them, are discussed below.

3.8.1 Incorrect selection of teachers for programme participation

Issue: The baseline survey data on teacher background characteristics show that about 16% of teachers in the treatment and control schools do not teach any of English, maths or science in primary grades, and it is thus unclear how they were selected into the programme (see Box 1 for the programme selection criteria), since TDP's in-service training component only covers these three subjects. It is uncertain to what extent this 16% of teachers would benefit from TDP's in-service training materials (or, if they are control teachers, how they could be deemed a valid counterfactual for treatment teachers).

Risk: The incorrect selection of 16% of teachers for programme participation for the roll-out of the TDP (and therefore also for the evaluation sample) means that the teacher sample size has been substantially reduced, which increases the MDE for programme impact on teacher effectiveness in its endline in 2018. Assuming that the design effect does not change due to attrition, the approximate increase in the standard error (and corresponding MDE) due to the smaller effective sample size is based on the square root of the ratio between the original sample size and the effective sample size, after removing the ineligible sample. In a case where the effective sample size is 84% of the original sample, the MDE would increase by a factor of about 1.09. However, assuming that the ineligible teachers are spread throughout the sample schools (so that entire schools do not drop out), this would actually reduce the design effect slightly, so it would be

reasonable to assume that the MDE would only increase by a factor of about 1.08. This is a rough estimate, but it gives an idea of the general magnitude.

If some replacement is used for endline surveys, whereby new teachers are enrolled for training to replace these 16% of teachers wrongly selected into the programme, this would not provide the same power for the longitudinal analysis as the original sample, but it should improve the precision of any cross-sectional estimates. Any replacement should be handled very carefully, to avoid introducing more bias. Alternatively, rather than dropping these teachers, it may be of interest to follow up at least a sub-sample of these cases in the follow-up survey, to see if the TDP training continues to be misaligned to the subjects they teach, or it may be possible that some of these teachers have switched to teaching the target subjects after all, in which case they could be brought back into the sample.

3.8.2 Potential contamination by other education programmes

Issue The TDP Phase 1 states coincide with those for other education programmes: in Jigawa, ESSPIN (also managed by Cambridge Education), and in Katsina and Zamfara, GEP3 (managed by the United Nations Children's Fund (UNICEF)). The TDP schools in Jigawa were existing ESSPIN clusters, many of which had not completed training and therefore will continue to receive ESSPIN interventions, in addition to the activities provided by TDP. TDP and GEP3 have made, and will continue to make, efforts to avoid implementing activities in the same schools in Katsina and Zamfara. However, this also needs to be negotiated with state partners, so it may not be possible in all cases to avoid contamination. In the baseline survey around 70% of TDP's treatment and control schools stated that they were in receipt of GEP/UNICEF support.

Risk: If other education programmes overlap with TDP implementation in TDP treatment schools, or implement education activities in TDP control schools, this would confound the evaluation's measurement of TDP impact with that of other programmes and it will not be possible to accurately measure TDP impact.

To address this problem will require accurate and timely information on the implementation of ESSPIN and GEP activities in TDP treatment and control schools. This will to be used in the impact analysis at endline (2018) to help account, to the greatest extent possible, for any contamination. The evaluation follow-up surveys will again seek this information from head teachers in interviews but such data can be notoriously unreliable and incomplete. Therefore the EDOREN evaluation team will also request TDP to provide lists of schools in which ESSPIN is implementing activities, and similarly for GEP3 from UNICEF.

3.8.3 Contamination of control schools by the programme

Issue During the qualitative fieldwork for the impact evaluation in June 2015 it was discovered that TDP in-service teacher training was being implemented in some control schools. In this observed instance, all control schools in one LGA were receiving TDP training, meaning that they had been contaminated and can no longer be used as control schools. At the same time, the TDP is not being implemented in any of the intended treatment schools in this same LGA, so that there is also a loss of treatment schools.

Risk As a result, the sample size has been reduced, which will affect the impact evaluation's ability to measure programme impact, both on teacher effectiveness and on pupil learning levels, at the endline.

At the moment, the extent of this type of contamination problem is not known. The implementation of the programme in schools that should not have been selected was discussed by the EDOREN evaluation team and TDP in June 2015 and it was agreed that TDP would provide the evaluation team with a complete list of schools in which the programme is being implemented, and, within these schools, which teachers are participating in the programme. This list has now been received by EDOREN. Once the evaluation team has had the opportunity to examine the TDP list it can be determined whether this was an isolated case or if the problem is more pervasive.

3.8.4 Attrition of sampled teachers due to transfer, death or inability to identify them

Issue: As discussed earlier, teachers who had been transferred, died or could not be identified were not included in the baseline survey. In order to retain the teacher sample size, after consultation between the EDOREN evaluation and TDP it was decided that the baseline data collectors would ask head teachers to name a replacement teacher using the selection criteria in Box 1, and these replacement teachers (from treatment schools) would be enrolled for the TDP in-service training. The EDOREN evaluation team shared this list of replaced teachers with TDP in January 2015 to enrol them into the programme in the treatment clusters for training (in September 2015). However, this means that, at the endline, these teachers will have received less training than other treatment teachers.

Risk: If TDP is successful in enrolling these replacement teachers in the programme there would be no reduction in the teacher sample size for this reason. However, if the TDP cannot do so the sample size would be reduced. The list provided by TDP to the evaluation team to examine the extent of contamination of control schools (see Section 3.8.3) will also be used to check to what extent these replacement teachers have been enrolled into the programme, and to ensure that only trained teachers are included in the follow-up surveys in 2018.

3.8.5 Attrition of sampled teachers due to invalid replacements

Issue: During the qualitative fieldwork for the impact evaluation in June 2015 it was also observed that there had been *ad hoc* replacements of some teachers (with other teachers in the same school) so these were not due to transfers, resignations or deaths.

Risk: This means that at the endline the survey will in some cases trace and interview, and assess and observe, lessons of teachers who should have participated in the programme but that have been incorrectly replaced and therefore have not received training. As a result, the evaluation would underestimate any programme impact on the selected indicators of teacher effectiveness and on pupil learning levels.

This problem has been discussed by the EDOREN evaluation team and TDP, and TDP will ensure that the teachers originally selected for training will receive the intended training, but with a delay. This means that this group of teachers will receive less in-service training compared to the intended two years of training.

3.8.6 New programme logframe

At the time of finalising this report, a new logframe for the programme, including new logframe indicators related to in-service training, had just been developed. The sample size calculations for the baseline survey are based on the old logframe and thus before the next survey round it will have to be determined what the new logframe implies for the evaluation design, and particularly for sample sizes and MDEs.

4 Qualitative research design

This section provides an overview of the objectives, framework, and method for the qualitative component of the mixed-methods impact evaluation of the TDP's in-service teacher training component. As part of the discussion of methods it sets out how rigour will be achieved through qualitative research; and our approach to sampling, training, fieldwork, analysis, and the limitations of the qualitative strand within the impact evaluation of the programme.

4.1 Why do we need qualitative research in an impact evaluation?

The key question this impact evaluation is expected to answer is *whether* and *by what mechanisms* TDP's in-service teacher training model improves teacher quality, and subsequently learning outcomes of pupils, in primary education in northern Nigeria. As discussed in Section 2 above, this impact evaluation applies a mixed-methods design, with qualitative and quantitative methods of evaluation. The quantitative school surveys (using several instruments aimed at gauging teacher effectiveness and learning outcomes) are designed to establish *whether* the in-service training has any impact, i.e. establishing causality between in-service training, and teacher pedagogical competence and subject knowledge (logframe outcomes) and pupil learning levels (logframe impact).

However, the quantitative evidence in itself does not provide insights into the underlying *mechanisms* driving causality. Given TDP, DFID Nigeria, and other stakeholders also need explanations for *how*, or the *mechanisms* by which, the in-service activities have had an effect on teachers and pupils, there is a need to accumulate lessons through explanatory and contextual analyses obtained by qualitative methods. As such, the reasons for conducting this companion qualitative research are to:

- support the quantitative survey results by providing a richer and more nuanced account of themes already covered by the quantitative survey (for example, *what* do teachers spend their time doing during class?);
- investigate themes which are not covered by quantitative surveys (especially those not amenable to quantification, for example *what* are head teachers' understandings of their responsibilities?);
- explore *how*, or the *mechanisms* through which, in-service activities have an effect on teachers and pupils, and therefore help to explain *why* variation exists between teachers or schools. (This will provide some indication of the likely transferability of quantitative findings and inform programme scale-up.); and
- provide preliminary answers to process-related evaluation questions related to the relevance, appropriateness, and implementation of the in-service training programme – particularly its cluster-based and peer-to-peer learning approaches – on the basis that TDP in-service training had already been implemented for approximately six months by the time the qualitative research took place.

The qualitative study uses case studies to explore the contextual circumstances that characterise the schools at various performance levels, in terms of learning outcomes and teacher effectiveness. Case study schools were selected based on the average TDNA scores of the schools, to ensure a wide distribution of schools. Such a case study approach of deviant cases is invaluable for generating findings for the 'how' and 'why' questions of the evaluation. Moreover, qualitative

methods are crucial for analysing the system-level context of each case study school because causal inferences do not occur independently from the context.

4.2 Using theory to understand variation and improve programme design

In the evaluation framework for this impact evaluation, the TDP evaluation team designed a basic theory of change (TOC) for the in-service training component of TDP, as presented in Section 2 of Volume I, which mapped out how programme inputs are expected to lead to impact (EDOREN, 2014). By setting out the suggested causal pathways through which change is assumed to take place, and the assumptions that can be expected to hold for these causal pathways to be activated, the TOC is a useful starting point for an exploratory analysis of *why* and *how* TDP achieves results, and how variation in *context* may explain variation in *results*. As such, this theory-based approach to qualitative research addressed the following three questions:

- First, based on the existing literature and researchers' own knowledge of the Nigerian education system, what are the expected assumptions that would need to hold in order for 'links' in the TOC (e.g. from output to outcome) to be bridged?

- Second, are these assumptions satisfied in the context of TDP and non-TDP schools (*partially addressed by the quantitative study*)?

- Finally, how and why do these assumptions influence whether TDP achieves its intended impact in the northern Nigerian context?

As such, the advantage of using the TOC as an entry point for this qualitative study was that it enabled the qualitative evaluation questions to be structured in such a way that they could be mapped onto the quantitative findings (i.e. in terms of output, outcome, intermediate impact); and, additionally, it enabled the researchers to start understanding how context explained variation in *baseline* findings.

A further advantage of carrying out a baseline qualitative study on the basis of the TOC will be that, in follow-up rounds of qualitative research, the evidence will enable researchers to suggest whether assumptions are valid *over time* for the selected sample of schools. This will help to address the challenge of relying on a snapshot assessment of the validity of programme assumptions only at follow-up stages to explain variation in programme impact, which will have been cumulative over the entire period over which the intervention has been implemented. Instead, the validity of assumptions can be assessed *over time* so that researchers can consider how assumptions at baseline may (partially) explain follow-up results for case study schools.

The existing TOC was turned into a qualitative evaluation matrix (Annex A.4) that identifies the expected assumptions that would need to hold in order for 'links' in the TOC to be bridged. These assumptions were then reformulated in terms of specific questions to be included in (more) structured instruments and observations. Emerging findings from the baseline qualitative research were then used to re-frame existing assumptions or add further assumptions that may not have been identified on the basis of the literature review and researchers' prior knowledge of the northern Nigerian context. The presentation of results in Volume I is thus categorised by these key assumptions.

4.3 Sampling

Data gathered through qualitative research cannot be quantified and aggregated in the same way as quantitative data, which means that randomised selection of a sample to achieve 'statistical representativeness' is rarely the most appropriate way to achieve methodological rigour in qualitative research. Nevertheless, a carefully designed approach to sampling remains an important way of achieving rigour in qualitative research. The sub-sections below detail the process of selecting a sample of schools, teachers and pupils for the qualitative study.

4.3.1 Selection of schools

Stratified purposive sampling was used to identify the final sample of schools used within the qualitative strand. This sample was nested within the contexts used for sampling in the quantitative strand. In particular, within each state (three states in total), the 56 treatment schools selected for inclusion in the quantitative study were first listed in descending order of the average of teachers' baseline scores in the teacher subject knowledge assessment (TDNA). These schools were then divided into three strata, representing the top 10% of schools ('high-performing'), the middle 10% ('typical schools', i.e. 45%–55%), and the bottom 10% of schools ('low-performing'). Control schools were removed as possible research sites. As the qualitative research is primarily concerned with the mechanisms by which the TDP in-service training influences school, teacher and pupil performance, and as this intervention was not given to the control schools, their presence in the sample would not add value. This decision was also motivated by time and budgetary constraints.

Two types of schools were removed from each shortlist: first, schools that were more than 90 minutes' travel from the state capital were removed for logistical and security reasons, because researchers could not visit these distant schools and return to the state capital on the same day before dark to carry out team debriefs. Second, schools that were located in LGAs that were deemed 'insecure' by security consultants were also removed from each shortlist. After filtering out these schools there were generally between two and four schools remaining within each strata. Finally, a school was randomly selected from within each strata's shortlist to study in-depth as part of the qualitative study. The process was repeated within each of the three TDP pilot states (Jigawa, Katsina and Zamfara), giving a total of nine schools across the three states. Table 10 below summarises this process.

This chosen approach to qualitative sampling is theoretically informed and pragmatically designed to generate responses from small numbers of individuals and groups that are representative (though not statistically representative) of groups relevant to the TDP, and that allow some identification of heterogeneous impact, within the constraints of budget, security and time. Therefore, the qualitative sample will also pick up extreme or deviant cases, as defined by teachers' TDNA performance in this case, as 'typical sampling' cases. This sampling process is expected to yield especially valuable evidence and interesting contrasts between schools, thereby allowing for comparability across and within cases.

Table 10 School selection process

Sampling frame for qualitative study	Filter		Filter		Filter	Qualitative sample per state	Repeat
56 treatment schools per state	Avg. TDNA score	Top 10% [6]	Accessibility	Top [~2–4]	Random selection	1 high-performing school	x 3 (for each state)
		Mid 10% [6]		Mid [~2–4]		1 typically-performing school	
		Bot 10% [6]		Bot [~2–4]		1 low-performing school	
Source: EDOREN evaluation team							

The limitations of this sampling approach are discussed in more detail in Section 4.7. In particular, it should be noted that this sampling method means that the school's performance category in respect of the TDNA is *relative to other schools in the state*. It is therefore possible (and indeed was the case) that typical or even low-performing schools in one state (e.g. Katsina) in fact had higher average TDNA scores than high-performing schools in another state (e.g. Jigawa) in a subject (e.g. maths). Similarly, it was also possible that schools with relatively high TDNA scores had worse pupil test scores than a school where teachers scored comparatively less in the TDNA. This was an important methodological consideration when conducting the analysis. Table 11 below summarises the key characteristics of the schools selected as part of the qualitative sample. The table does not include the LGA in which each school is situated, in order to protect the anonymity of sources at the local government level. More in-depth descriptions of schools are provided in Annex N below.

Table 11 Select school-level information for sampled schools

State	Performance category	English TDNA (%)	Maths TDNA (%)	English literacy pupil test (%)	Numeracy pupil test (%)	# teachers	# pupils	Average class size
Katsina	High	39	88	14	30	17	978	39
Katsina	Typical	20	47	33	47	19	1600	33
Katsina	Low	6	39	25	22	13	1010	68
Jigawa	High	52	61	18	26	14	980	111
Jigawa	Typical	28	37	13	15	8	350	36
Jigawa	Low	2	29	16	24	5	266	20
Zamfara	High	30	50	18	42	27	1125	21
Zamfara	Typical	23	32	13	25	5	256	65
Zamfara	Low	12	33	21	32	9	299	35

Source: TDP in-service training impact evaluation baseline survey (2014); TDNA and pupil test scores represent marks scored (raw) as a percentage of total marks; # teachers = number of teachers employed (Grade 1–6); # pupils = number of Grade 1–6 pupils registered

4.3.2 Selection of TDP and non-TDP teachers

All four TDP teachers (including the head teacher) in each school were included as part of the qualitative fieldwork. Non-TDP teachers (from whom quantitative data were not collected in treatment schools) were sampled by the subjects they teach, i.e. they must teach at least one of the three TDP subjects (English, maths and science). Where possible, a gender and age balance was also sought in the sample of non-TDP teachers within each school.

4.3.3 Selection of case study teachers

One teacher was purposively sampled from among the TDP teachers for an in-depth case study in each school. These teachers were selected in advance of visiting the school on the basis of baseline quantitative data. In each state, research teams coordinated to ensure that case study teacher profiles were varied across schools in terms of age, gender, and subjects taught, based on the quantitative baseline survey data at hand. This enabled researchers to compare the experiences of teachers at different points in their careers, and potentially across teachers with varying personal characteristics, and teaching obligations.

4.3.4 Selection of pupils

Where possible, the intention was to carry out the FGDs separately for boys and girls within the qualitative strand on the same set of eight pupils who had been randomly selected from Grade 3 classes and administered the learning assessment as part of the quantitative baseline survey. In almost all sampled schools, however, not all of these pupils (who were tested in the baseline survey) were available to populate the baseline qualitative sample on the days of the survey. In such cases, additional pupils were included to 'top up' the FGD sample. These were randomly

selected from Grade 3 classrooms, though in a few cases, where no more Grade 3 pupils were available, they were also randomly selected from other grades.

4.4 Data collection tools

4.4.1 Structured and unstructured methodologies

The qualitative part of the impact evaluation made use of three main research techniques or instruments –KIIs, FGDs, and lesson or cluster meeting observations. All of the KIIs, FGDs and observations utilised structured and unstructured methodologies. Structured methods allowed for the efficient probing of pre-specified assumptions in the TOC, while unstructured methodologies allowed for unanticipated or context-specific information to be captured, as well as new assumptions to be developed and explored. Observations tended to be more unstructured, though they were guided by process-based evaluation questions (in the case of cluster meetings) or guided by pre-determined questions that facilitated an overall assessment of the extent to which teaching practice was characterised by behaviours associated with child-centred learning (in the case of lesson observations). Table 12 summarises the types of instruments that were used to collect information from different participants in the qualitative research.

Table 12 Instruments administered for each participant group

Instrument	Head teachers	TDP teachers	TDP case study teacher	Non-TDP teachers	Pupils (girls)	Pupils (boys)	LGEA	SUBEB	TDP TF	TDP cluster meetings	TDP, DFID, ESSPIN and GEP programme staff
KIIs	x	x†	x	x†			x	x	x		x
FGDs					x	x					
Lesson observation			x							x	

† These were group interviews that included some interactive elements more typical of FGDs.

4.4.2 Design of the KII, FGD and observation instruments

The design of the KII, FGD and observation instruments was based on the questions identified in the process of developing the qualitative evaluation matrix (Annex A.4). These questions were identified to: (i) further investigate unexpected or hard-to-interpret quantitative findings; (ii) assess whether expected assumptions in the TDP TOC were likely to hold in each case; (iii) identify the mechanisms by which context influences schooling outcomes; and (iv) provide some preliminary answers to process-related evaluation questions pertaining to the relevance, appropriateness, and implementation of the in-service training. Answers to these questions at baseline will help inform the design of the follow-up rounds of the quantitative survey and qualitative component. Each instrument focused on the core areas to probe that were most relevant to the group of individuals participating in that instrument. Table 13 below provides an

overview of the key components of the TOC, the areas explored by each instrument, and the source(s) of information for each of these areas.

These core areas probed were grouped into broad themes, around which the questions in the KIIs were organised. On the one hand, these themes were used to ask participants questions related to these themes and to structure debrief discussions among researchers. On the other hand, the flexibility of these themes also allowed us to capture unexpected responses that did not correspond strictly to the questions included in the qualitative evaluation matrix. This semi-structured approach provided a degree of standardisation, whilst allowing the researchers enough flexibility to pick up on interesting themes, topics and concerns as they emerged during interviews and discussions.

Written consent to participate in discussions, and to record responses, was sought from each adult participant before commencing the discussion, and from the head teachers on behalf of the pupils. Verbal assent was sought from pupils (Section 5 details the consent provided, and the ethics of research, for the children in the sample). Each instrument then started with a tailored ice-breaking discussion or activity, to put participants at their ease. For KIIs, this generally included a brief discussion of the participants' background and offering participants an opportunity to ask questions to the interviewer. For FGDs, ice-breaking involved the facilitators and pupils singing a traditional Hausa song, with accompanying gestures. All instruments were conducted in Hausa (the lingua franca of northern Nigeria).

Table 13 Core areas probed and sources of information from the qualitative evaluation matrix

Area of impact	Core areas to probe (including change over time for each)	Source of information
Intermediate impact: Improved teacher effectiveness in classroom	• What **motivates** children? • Why don't children **attend** lessons at school? • How do pupils **behave** during lessons? • How might other obligations outside of school (**caring** for family members, household **chores**, **income**-generating activities) limit pupils' school attendance or time spent on learning outside of school? • What support do **parents** provide to pupils? • Why are learning outcomes often lower amongst **girls** and children from the **poorest households**? • How do (head) teachers **monitor** pupils' learning outcomes on an ongoing basis, and how does this influence teacher and management practices? • How do teachers mitigate the potential negative impact of **language** differences?	• Pupil FGD • Head teacher KII • Teacher FGD • Informal discussion with community members/SBMC
Outcome 1: Improved head teacher SLM	• What **motivates** head teachers? • What **incentive structure** do head teachers face, and how does this influence their decision-making? • How do the **state government/LGA** and **local community** hold head teachers to account? • How do head teachers communicate their view of what constitutes an 'ideal' teacher to their staff? • How do head teachers **plan** for the future and how do they request the **resources** to implement this plan? • How have new leadership and management techniques been **received** by teachers and community members? Has here been any **tension** between school changes and local values, and how has this been managed? • How has TDP influenced the **relationships** between teachers receiving training and other teachers?	• Head teacher KII • LGEA/SUBEB KII • Informal discussion with community members/SBMC
Outcome 2: Improved teacher subject content knowledge	• What **motivates** teachers? Why are Nigeria Certificate in Education- (NCE-) qualified teachers more motivated than unqualified ones? • Why do teachers not attend school and how do they mitigate the negative impact of **out-of-school obligations** on their school attendance (if at all)? • Why do teachers curtail **lesson duration** and how do head teachers address	• Teacher FGD • Teacher case study + photo exercise • TF KII • Head teacher KII

Area of impact	Core areas to probe (including change over time for each)	Source of information
	this (if at all)? • What **incentive structure** do teachers face, and how does this influence their teaching practice? • How do the head teacher**, school inspectors** and **pupils' parents** hold teachers to account? • Why does a lack of **school infrastructure** (classrooms, furniture, sanitation) present a barrier to the application of new subject knowledge, and how do teachers mitigate the impact of this limitation? • Why does a lack of **classroom materials** (textbooks, aids) present a barrier to new subject knowledge, and how do teachers mitigate the impact of this limitation?	• LGEA/SUBEB KII • Informal discussion with community members/SBMC
Outcome 3: Improved teacher pedagogical knowledge	• As Outcome 2 above, and: • What do teachers feel about the appropriateness and flexibility of the **curriculum**?	• Teacher FGD • Teacher case study + photo exercise • TF KII • Head teacher KII • LGEA/SUBEB KII • Informal discussion with community members/SBMC
Output 1: Collaboration and partnership	• Why do teachers (not) **attend** cluster meetings? • What do teachers **discuss** during cluster meetings? • How **free and comfortable** do teachers feel in regard to speaking up during cluster meetings? • How do **gender** or other factors influence participation in cluster meetings? • To what extent do teachers share knowledge in **informal, less structured contexts**?	• Teacher FGD • TF KII / cluster meeting observation • Head teacher KII
Output 2: Training and support	• What **motivates** (head) teachers? • What **incentive structure** do (head) teachers face, and how does this influence their willingness to learn? • What do (head) teachers believe constitutes **an 'effective' (head) teacher**? • How do (head) teachers receive **feedback** on their learning that enables them to know when they are learning correctly? What role do head teachers, TFs, and the learning resources themselves play in this? • Why do teachers not participate in training opportunities and how do they mitigate the negative impact of **out-of-school obligations** on the time they spend training (if at all)?	• Teacher FGD • TF KII / cluster meeting observation • Head teacher KII
Output 3: Development of materials	• To what extent do head teachers **encourage** use of TDP teaching materials? • How do teachers **feel** about the new teaching materials? Do they make **links** between newly acquired pedagogical knowledge and new teaching materials?	• Teacher FGD • Head teacher KII
Output 4: Technology use and management	• How do teachers **feel** about the audio-visual resources? Do they make **links** between audio-visual resources and new teaching materials? • To what extent do teachers feel that the audio-visual resources are **delivered** in a way that improves their learning? • How do (head) teachers receive **feedback** on their learning that enables them to know when they are learning correctly? • Why do teachers not use the 'teacher in the pocket' resource more and how do they mitigate the negative impact of **out-of-school obligations** on the time they spend using it (if at all)?	• Teacher FGD • Head teacher KII
Outcome 5: In-Service Education Training (INSET) implementation	• How do teachers **feel** about the INSET? Do they make **links** between INSET and their role as a teacher? • To what extent do teachers feel that INSET is **delivered** in a way that improves their learning? • How do (head) teachers receive **feedback** on their learning that enables them to know when they are learning correctly? • Why do teachers not use the INSET resources more and how do they mitigate the negative impact of **out-of-school obligations** on the time they spend using it (if at all)?	• Teacher FGD • TF KII • Head teacher KII

Area of impact	Core areas to probe (including change over time for each)	Source of information
Process 1: Collaboration and partnership	• To what extent do teachers feel that cluster meetings are scheduled at convenient times and locations that minimise disruption to other obligations and cost of attendance?	• TDP KII • Teacher FGD • TF KII / cluster meeting observation
Process 2: Training and support	• What are the roles and responsibilities of TFs? • What challenges do they face, if any, in delivering their role as TFs?	• TDP KII • Teacher FGD • TF KII
Process 3: Development of materials	• Have teachers received teaching materials? If not, why not?	• TDP KII • Teacher FGD
Process 4: Technology use and management	• Have teachers received 'teacher in the pocket' resources? If not, why not?	• TDP KII • Teacher FGD • TF KII
Process 5: INSET implementation	• Have teachers received INSET? If not, why not?	• TDP KII • Teacher FGD • TF KII
Source: TDP in-service baseline qualitative survey evaluation matrix (Annex A.4)		

The draft instruments were tested as part of two sets of pilots. The first set was conducted in three schools in Kaduna State, given its proximity to Abuja, where the week-long training of national researchers was taking place. This pilot was mostly used as an opportunity for researchers to improve their probing skills and to familiarise themselves with the tools. However, revisions to the structure of some tools were made to ensure they could be completed within the time allocated. A second set of pilots was conducted in three schools in Katsina State. The focus of these was to revise any questions and translations of tools to ensure they were easily understood by participants, and to practice preliminary in-field analysis and debriefs among researchers.

4.4.3 Description of the head teacher KIIs

The KIIs with head teachers covered a broad range of themes relating to career history; personal motivation; feelings of being valued; the teacher recruitment process; personal and professional characteristics of effective teachers; identification of effective teachers; opportunities for improving subject knowledge and teaching technique; leadership and relationships with teachers; reporting lines; school inspections; interaction with community and parents; school resources; out-of-school obligations; advice they would give to new head teachers; and process-related questions regarding the implementation of TDP.

4.4.4 Description of the TDP and non-TDP group KIIs with teachers

The group KIIs with TDP and non-TDP teachers covered themes relating to career history; personal motivation; feelings of being valued; teacher recruitment and career progression; curriculum; teacher training; language; advice they would give to new teachers; lesson observations; relationship with (non-)TDP teachers and head teachers; interaction with parents; and out-of-school obligations.

The group KIIs included a 10–15-minute constraint-ranking activity, in which teachers were asked to arrange 12 constraints, printed and given to them on slips of paper, in order of their perceived negative impact on teachers' ability to teach effectively. The primary purpose of this exercise was to generate discussion between participants about *why* certain constraints were seen as

particularly binding. Participants were encouraged to provide examples to justify their decisions. The constraint-ranking exercise was not designed to generate findings regarding perceived constraints that were statistically representative of teachers in the TDP intervention clusters more generally: the order of constraints was therefore of secondary interest to researchers. However, the ranking exercise was useful to validate researchers' interpretations of participants' responses given during the rest of the KII. This was particularly useful in a context in which teachers would sometimes complain about nearly every aspect of being a teacher during the majority of the KII, and there was a risk of placing too much burden on researchers' own interpretation of participants' views when considering the scale of constraints *relative to each other*.

Table 14 Constraint categories for ranking exercise

Examples of constraint categories for ranking exercise provided to teachers		
Lack of leadership from head teacher	High pupil absenteeism	High teacher absenteeism
Inadequate school infrastructure[33]	Poor pupil behaviour	Low teacher motivation
Inadequate classroom resources[34]	Untalented pupils	Poor teacher subject knowledge
Inappropriate curriculum / textbooks	Lack of support from parents	Poor teaching technique
Source: TDP and non-TDP teachers' KII instruments		

The group KIIs with TDP teachers also included questions relating to their experience of receiving the first six months of training and support from TDP. This was followed by a short exercise in which participants were asked to indicate which of the constraints listed above they believed TDP might influence, and whether this influence was positive or negative. The purpose of this exercise was to encourage participants to articulate any unanticipated mechanisms through which TDP may achieve (or undermine the achievement of) its desired impact.

4.4.5 Description of the TDP teacher case study

One of the teachers receiving TDP training (except the head teacher) was selected for an in-depth case study from within each of the case study schools (nine in total). The case study comprised three elements: a lesson observation, a photo diary, and a KII. The lesson observation involved drawing a sketch of the classroom (including marking the furniture, blackboard or other noteworthy features), counting the number of pupils in a class, mapping the movement of the teacher in the classroom, and considering questions that enabled an assessment of the extent to which teaching techniques followed principles of child-centred learning.

Research teams spent three days in each school. Case study teachers were given a digital camera on the first day in each school and asked to take photos that would help researchers understand the life of a public primary school teacher in parts of northern Nigeria. Teachers were encouraged to take as many photos as they liked, and then to select 10 photos to be discussed on the last (third) day as part of a KII. The objective of the photo study was to empower TDP beneficiary teachers to set the agenda of a discussion by choosing what to focus on in their photos. The aim was to elicit episodic storytelling in which each teacher gave his or her own interpretation of what the photo showed by describing the events surrounding the photo. The researcher facilitated discussion using prompts and, where appropriate, encouraged respondents to think about how the events in the photos related to their effectiveness as teachers (or the effectiveness of TDP).

[33] Includes electricity, quality of building, furniture, blackboards, toilets and water supply.
[34] Includes textbooks, exercise books, pens, pencils, posters, audio speakers and batteries etc.

Such visual documentary data may enhance the ability of teachers to communicate their perspectives, which may not necessarily be captured by verbal communication.

Finally, the case study teacher KII also allowed researchers to follow up on points mentioned by other informants or previously in the group interview in a more confidential setting where individual teachers were sometimes more willing to express themselves openly. Case study teachers were also presented with vignettes (short unfinished stories depicting a concrete real-life scenario) and asked to describe what they would do in such a scenario. The aim was to stimulate teachers' thinking about particular topics, and to make it easier for them to articulate their thoughts in relation to a class of situations. Topics covered by the vignettes touched on teacher motivation, pedagogy, accountability, support, relationships with other teachers, and community relations. Not all vignettes were presented to all case study teachers due to time constraints. Instead, the most appropriate vignettes were selected on the basis of dynamics discussed previously in the teacher case study or other instruments in the school.

4.4.6 Description of the pupil FGDs

The two FGDs with Grade 3 pupils covered themes relating to: pupils' learning experiences; relationships with teachers; parental support with studies; home environment; reasons preventing pupils from attending school; gender differences between how girls and boys perceive schooling and learning; and gender differences in the schooling experience.

Pupils completed three activities that enabled researchers to gain insights into these themes. First, pupils were asked in an unstructured way about their day at school, using pupils' exercise books (where available) as a prompt for discussion of pedagogy. Second, researchers drew an outline of a happy girl/boy on A1 paper and asked pupils to tell them what characteristics this happy girl/boy had. Where responses were not forthcoming, pupils' attention was guided towards describing the happy boy's/girl's material conditions, family background, daily routine and experience of school. In order to understand pupils' priorities or aspirations, pupils were asked to imagine they were a popular folklore character with special magic powers to make dreams come true, and then asked what they would change about their school. Third, pupils were given coloured crayons and asked to draw the activities that they carry out when they are not at school. These drawings were discussed as part of a group, and the pupils were probed to investigate how often the activity was carried out, whether pupils enjoy it, whether their friends do the same, and whether it affects their schooling.

4.4.7 Description of TDP TF KIIs and cluster meeting observations

The KIIs with TDP TF and cluster meeting observations were a key component of the preliminary process evaluation of TDP's implementation. They covered themes relating to: career background; the TF recruitment process; roles and responsibilities; challenges; the resources available; identification of effective teachers; problems faced by teachers, and solutions; TDP cluster meetings; and TDP school support visits. TDP TFs were also presented with a vignette tailored to their role, which aimed to assess how they would diagnose and address low teacher subject knowledge. The choice of this vignette was informed by the quantitative baseline survey, where evidence showed widespread low teacher subject knowledge in all three subjects.

The observation of TDP cluster meetings was structured in terms of key hypotheses that were subject to confirmatory analysis. These hypotheses were grouped into themes relating to: experience sharing; facilitation style; materials; administration; and language.

4.4.8 Description of the LGEA and SUBEB KIIs

The KIIs with LGEA and SUBEB officials covered themes relating to career background; roles and responsibilities; interaction with others in the education system; understanding of challenges presented by gender and income differences; teacher training; TDP cluster meetings; TDP school support visits; perceptions of teachers; policies and practices to support teachers; perceptions of head teachers and SLM; and policies and practices to support head teachers.

4.4.9 Description of the TDP, DFID, ESSPIN and GEP KIIs

Programme staff were interviewed from the TDP management teams in Abuja and state offices. These KIIs covered themes relating to: the rationale behind the TDP teacher training model; the design process; teacher recruitment; specific challenges teachers face, as identified up to that point in the fieldwork; teacher motivation; understanding of challenges presented by gender and income differences; mechanisms for beneficiaries to provide feedback on TDP; selection and training of TFs; cross-learning between programmes; and engagement with government.

Programme staff from DFID, ESSPIN (also managed by Cambridge Education) and GEP (managed by UNICEF) were also interviewed to understand their experience of, and input into, the TDP design process; the coherence of DFID-funded education interventions in northern Nigeria; and the extent to which knowledge and lessons were shared. These instruments covered similar themes to the KIIs with TDP programme staff, and were also used to validate emerging findings from the qualitative fieldwork.

4.5 Fieldwork

The fieldwork for the qualitative research took place over three weeks in June 2015, immediately following one week of training for the local researchers in Abuja (Table 15). The federal and gubernatorial elections in Nigeria in March 2015 meant qualitative research fieldwork could not be conducted immediately after the qualitative baseline survey, which ensured that findings from the quantitative survey could be fully factored into the design of the qualitative data collection instruments.

Table 15 Timeline for qualitative data collection, cleaning and analysis

Activities	Dates	Location
Finalisation of concept note, tools, fieldwork guide and recruitment of local researchers	April/May 2015	Oxford/Abuja
Training for fieldwork manager and local researchers (including one-day pilot in Kaduna)	1–6 June 2015	Abuja
Field pilot	8–9 June 2015	Katsina

Activities	Dates	Location
Data collection: Katsina	10–12 June 2015	Various LGAs
Data collection: Jigawa (including observing cluster meeting observations)	15–17 June 2015	Various LGAs
Data collection: Zamfara (including observing cluster meeting observation)	18–19 and 22 June 2015	Various LGAs
KIIs with DFID, TDP, ESSPIN and GEP3	24–26 June 2015	Abuja
Transcription and translation of interviews	July 2015	Abuja
Data cleaning, analysis and report writing	August–September 2015	Oxford
Presentation of preliminary qualitative findings to DFID, TDP and TDP annual review team	October 2015	Abuja
Submission of quantitative findings integrated into a mixed-methods report to TDP evaluation steering committee	November 2015	Abuja
TDP evaluation steering committee meeting to discuss feedback and future of the evaluation	January 2016	Abuja
Receipt of detailed feedback from DFID and TDP	February–March 2016	Abuja
Finalisation and submission of mixed-methods report to DFID for SEQAS review	April 2016	Oxford/Abuja
Source: EDOREN authors		

4.5.1 The fieldwork teams

The qualitative fieldwork team included both international and national researchers. International researchers were selected on the basis of strong qualitative design, data collection and analysis skills, specialism in education policy and practice, and familiarity with the northern Nigerian context. National researchers were selected on the basis of previous research experience, formal training, familiarity with the northern Nigerian education system, fluency in Hausa, and flexibility to adapt to the principles of qualitative research. In most cases, the international and national consultants had conducted previous EDOREN-managed research on the northern Nigerian education system, either as part of the TDP quantitative baseline research, the ESSPIN impact evaluation or the EDOREN Teacher Management Study (Watts and Allsop, 2015).

A one-week training session was held in Abuja immediately prior to fieldwork to ensure researchers were familiar with the principles of qualitative research (including ethical considerations), the logic underpinning the TDP intervention and the design of the various instruments, and the practice of note-taking. Although the researchers had previous experience

with research, each organisation has its own standards for what constitutes quality research and the team was therefore trained to meet EDOREN standards. The training was a two-way process, in which feedback from national researchers was used to improve international researchers' sensitivity to local context and revise the design of instruments.

At the start of fieldwork the team was divided into three sub-teams each comprising of one international researcher and two national researchers. These were constituted to ensure gender and age diversity within teams. For each instrument, national researchers were assigned the role of interviewer/facilitator or note-taker, depending on the profile (especially gender) of instrument participants. International researchers acted as critical observers of non-verbal data (e.g. context, body language, school infrastructure) and supported national researchers when they sought guidance on how best to probe participants' responses. During the training and initial stages of fieldwork, these sub-teams were supported by an international expert in qualitative evaluation methods, who roamed between the groups to ensure consistency in approach and provide expert technical guidance.

Each sub-team spent three days in one of the three schools sampled in each state. This had the advantage of teams being close enough to each other to hold an inter-team debrief at the end of each day of data collection. Across the three states, each sub-team was assigned to one high-performing, one typical, and one low-performing school. Ensuring sub-teams were exposed to schools with a range of performance categories and facilitating inter-team dialogue were important steps that were taken in order to improve the trustworthiness – and therefore rigour – of the research process. The fact that all researchers were exposed to schools from a range of performance categories enabled all researchers to participate in discussions about the potential drivers of 'low' and 'high' performance, and therefore decreased dependence on a small set of researchers' own experiences. The process of daily inter-team debriefs was important in increasing the confirmability of results (i.e. decreasing the extent to which biases, motivations, interest or perspectives of the inquirer influenced interpretation), since it allowed researchers in one team to 'sense-check' interpretations of responses with researchers in another team.

4.5.2 Sequencing of instruments in the field

On arrival in each of the three states, introductions were made to senior SUBEB officials, who in turn facilitated introductions to the LGEAs and Education Secretaries, and head teachers of schools sampled. Head teachers were informed a day in advance of the arrival of research teams, which may have influenced teacher attendance and behaviour in school. School and lesson attendance of teachers reported during the qualitative study is therefore likely to offer a more positive impression than would usually be the case. However, neither head teachers nor teachers were informed of the topics to be discussed as part of the qualitative research, so it is unlikely that answers were prepared in advance.

KIIs with head teachers were conducted first on arrival at each school: this played an important role in putting head teachers at ease regarding the kinds of questions researchers would be asking in their school. The support of head teachers was enlisted in mobilising KII and FGD participants, though the head teachers' discretion was limited by the pre-selection of TDP teachers, the case study teacher, and the child participants in the schools. Head teachers were able to exercise most discretion in the selection of non-TDP teachers, since these were not pre-selected by the research team. However, research teams expressed preferences for non-TDP teachers that taught TDP subjects (English, maths and science) and encouraged gender and age diversity in the sample. In

each case, consent was sought from participants before questioning. It was also made clear to participants that they were free to end their participation in the interview or discussion at any time.

Case study teachers were approached on the first day in order to provide them with the digital camera for the photo diary instrument. Lesson observations were also carried out on the first day, where possible, so that case study teachers did not have time to prepare their lesson plans and to show researchers an atypical lesson. Enquiries were made with the case study teacher on the second day to ensure they had not encountered any difficulties with the digital cameras. The case study KIIs were held on the third day to allow case study teachers sufficient time to take photos.

The timing of the remaining tools was flexible and adjusted according to the timetable of each school, in order to minimise keeping teachers or pupils out of classrooms. Interviews with LGEA officials were generally left until the third day in each school to allow for preliminary analysis of the school-level qualitative data to inform the focus of the unstructured component of the LGEA KIIs. Where this was not possible, LGEA KIIs were held on the second day.

In a few cases the sampling was adversely affected by teacher and pupil absenteeism. However, the impact of this was minimised by spending three days in each school, so that informants missing on one day were requested by the head teacher to attend school in order to participate in the research the following day if possible. This approach was not taken with the pupils FGDs, where missing pre-selected pupils were replaced with other available pupils randomly selected from Grade 3.

The KIIs with SUBEB officials were left until the last day spent in each state. As with the LGEA interviews, this allowed for preliminary analysis of the qualitative data from all three schools – and emerging state-level patterns – to inform the focus of the unstructured component of the SUBEB KIIs. The KIIs with programme staff from TDP, DFID, ESSPIN and GEP were conducted in Abuja (or in some cases remotely from the UK) after state-level data collection had taken place – again, so that emerging findings could inform the approach taken in the unstructured component of these interviews.

4.5.3 Debriefs and team checks in the field

Daily debriefs within sub-teams were carried out after data collection each day. Since interviews were conducted in Hausa without live translation of every detail into English, these were particularly important opportunities for international and national researchers to clarify interview responses and identify emerging findings at the school level. Debriefs also allowed for the development of hypotheses that needed to be investigated, both through collection of additional data during subsequent visits and through confirmatory analysis. Finally, the process of discussion allowed for a critical cross-examination of researchers' interpretations of qualitative data, which further improved the trustworthiness of the analysis.

Debriefs between sub-teams were held following each within-team debrief. Each sub-team shared summaries of their day's findings and presented the hypotheses that they would investigate in subsequent days. This iterative approach to research made it possible to identify whether similar dynamics existed across schools and whether there was variation, and if so to suggest factors that might explain this variation. The discussion was structured in terms of the themes used to group questions across the instruments. International researchers maintained a fieldwork journal that summarised these debriefs.

The performance categories of each school were revealed to national researchers on the evening of the second day in each school. Prior to this, national researchers had not been informed as to whether the school under investigation was considered high-performing, typical, or low-performing. This facilitated a discussion about whether the ranking was surprising given the dynamics that had been identified in each school. Where such surprises were encountered, attempts were made on the third day to investigate why the school defied previous expectations: in particular, whether this was a result of the way school performance was measured (i.e. on the basis of average TDNA scores rather than learning outcomes), whether some other hitherto undetected dynamic explained the deviation from the expected performance category, or whether researchers' own personal biases had rendered previous interpretation of data incorrect.

4.5.4 Data treatment and collection

Researchers were provided with note-taking forms tailored to each instrument. The aim of these was not for researchers to take down a verbatim record of participants' responses, but rather to record enough data to facilitate the daily team debriefs. Notes from debriefs were typed at the end of each day and provided the basis for preliminary synthesis of findings.

Audio recordings of instruments were sent to transcribers following completion of the fieldwork. The required ethical standards regarding treatment of data – in particular, anonymity and data security – were carefully explained when sending the data outside the fieldwork team. Transcription was a time-consuming and extremely challenging process, especially given the levels of background noise in some recordings. In a small number of instances, the responses to questions were not audible from recordings. In these cases, the researchers resorted to the note-taking forms to fill gaps where possible. These transcriptions were extremely useful in allowing for detailed analysis of responses by international researchers on their return to the UK.

The lesson observation sketches, pupil FGD drawings, and consent forms were also safely stored for further analysis.

4.6 Analysis

4.6.1 Approach to analysis

The approach to analysing the qualitative data was based on thematic analysis. Thematic analysis is an inductive approach to research that requires more involvement and interpretation from the researcher. Thematic analysis rejects a quantitative approach to analysing qualitative data (such as frequency or cluster analysis) and instead focuses on interpretation of the stories and experiences shared by participants in order to identify and examine themes in as rigorous a way as possible. The analysis included elements of both confirmatory and exploratory approaches. Insofar as the design of the qualitative evaluation was informed by the TDP TOC, analysis was guided by the investigation of pre-specified hypotheses (i.e. it was confirmatory). However, insofar as these hypotheses were revised, broken down into sub-hypotheses and reconstituted as part of an iterative process, the analysis was principally driven by the content (i.e. it was exploratory).[35]

[35] The research team had originally intended to use the context-mechanism-outcome configuration that characterises realist evaluation as the main data analysis framework. However, this framework was deemed inappropriate for the baseline study, and was therefore revised, due to: the difficulty of achieving a shared understanding of how to interpret the same data among the research team, and challenges in identifying the appropriate 'outcome' to be analysed during the baseline study. Realist

4.6.2 Managing analysis in parallel

Data analysis was carried out in NVivo, Version 10, and undertaken concurrently by international researchers responsible for analysis of particular areas (pupils, teachers, SLM, the process of TDP implementation). An initial master file was created and populated with transcripts (sources) and an initial list of themes (nodes). All sources were initially coded at nodes containing metadata (attributes) related to the school and participants' role. Each researcher copied the master file and undertook coding and analysis in parallel on their local copies. These local copies were consolidated at regular intervals by importing them into the master file, and researchers were then re-issued with the master file. Given the common sources and initial starting set of nodes provided by the master file, there were no reported data corruption or duplication issues associated with merging the local files. Researchers held frequent meetings to discuss emerging findings and approaches to interpretation to ensure a coordinated and coherent approach to analysis.

Researchers used techniques in addition to theme identification to facilitate interpretation. These included word searches, coding queries, and matrix coding queries to understand both how themes related to each other (e.g. to what extent is teacher absenteeism spoken about in the context of low teacher motivation?) and to compare the same theme across different schools and role groups (e.g. do teachers and head teachers share a common understanding of X?).

To ensure that data analysis was rigorous, a range of principles and strategies were implemented during analysis. The specific tools were structural coherence and systematic data analysis, disciplined subjectivity, team data analysis, and triangulation. The main focus for the data analysis stage was to minimise a single researcher bias, to ensure transparency of data reduction and coding process and to develop a clear trail of the data analysis process such that the reader is able to see how findings emerged and conclusions were made. Thus, for every assumption tested, researchers would ascertain whether the assumption was met ('Yes' or 'No' or 'Mixed'/'Inconclusive') and then indicate the strength of evidence on the basis of which this decision was made. Clear standards were agreed between researchers for assessing the strength of evidence applied to conclusions about whether assumptions in the TOC held. These are summarised in Table 16 below.

evaluations are designed to be conducted *after* an intervention has been rolled out and aim to analyse *how* and *why* the intervention achieved the results it did. However, during baseline studies the 'outcome' cannot be defined in to the intervention's results.

Table 16 Definitions of strength of evidence used by researchers

Strength of evidence	Definition
Strong	Evidence is *both* frequently explicitly mentioned in interviews *and* supported by researchers' own inferences from interviews, context and background literature.
Mixed	Evidence is *either* (a) *both* explicitly mentioned in a minority of interviews *and* supported by researchers' own inferences from interviews, context and background literature *or* (b) frequently explicitly mentioned in interviews *but not necessarily* supported by researchers' own inferences from interviews, context and background literature. OR There is strong/moderate/weak evidence that an assumption holds in at least two cases but there is strong/moderate/weak evidence that the same assumption does not hold in at least another two cases.
Weak	Evidence is *either* explicitly mentioned in a minority of interviews *or* supported *only* by researchers' own inferences from interviews, context and background literature.

The combination of thematic analysis, use of techniques, and requirement that strength of evidence be explicitly supported with textual references means that the overall approach shares similar characteristics to applied thematic analysis (Guest, MacQueen, and Namey, 2011).

4.6.3 Final assumptions investigated

A large part of the qualitative study was concerned with investigating whether the assumptions that would need to hold in order for the TDP TOC as a whole to hold did in fact hold at baseline. The assumptions that were specified before the fieldwork were revised as part of the exploratory analysis before being subjected to confirmatory analysis. Table 17 below shows the final assumptions investigated during the qualitative analysis. These were broken down into sub-assumptions to facilitate easier interpretation by readers.[36]

[36] Note that some assumptions identified in the qualitative evaluation matrix were not deeply investigated at baseline due to resource constraints (e.g. 'parental encouragement of pupils to learn by reinforcing positive behaviour and learning outcomes' which bridges the link between improved teacher effectiveness (intermediate impact) and improved pupil learning outcomes (impact)).

Table 17 Assumptions investigated during qualitative analysis

#	Assumption	Relevant links in TOC
1	- Pupils **attend** lessons at school - Pupils have sufficient **nutrition** and rest to facilitate concentration during lessons - Pupils have **time** outside of school hours to reinforce learning (e.g. through discussion and homework) - **Parents encourage pupils** to learn by reinforcing positive behaviour and learning outcomes outside school - **Social norms** and **power relations** do not prevent pupils from engaging with new teaching techniques - **Gender** and **economic inequalities** of pupils are identified and their impact on learning outcomes is mitigated by teachers - (Head) teachers have **feedback mechanisms** that enable them to identify improved learning outcomes and reinforce positive teaching and management practices - Differences between instruction **language** and a pupil's primary language do not prevent transmission of knowledge	• Intermediate Impact ('improved teacher effectiveness') to Impact ('Improved learning of English, maths and science and technology for cohorts taught by selected teachers in TDP schools in 2014–2019')
2a	• **Teachers are motivated or incentivised** to attend school and lessons regularly, to try to improve their teaching, to take part in learning opportunities and to apply new knowledge when they get it	• TDP training activities and outputs to Outcome 2 ('Improved teacher subject content knowledge) and Outcome 3 ('Improved teacher pedagogical knowledge) • Outcome 2 to Intermediate Impact ('Improved teacher effectiveness in classroom') • Outcome 2 and Outcome 3 to Intermediate Impact
2b	• Teachers have the **foundational subject knowledge** to be able to understand training and curriculum materials	• Outcome 2 ('Improved teacher subject content knowledge) to Intermediate Impact ('Improved teacher effectiveness in classroom')
2c	• Teachers have a **foundation of pedagogic knowledge** to be able to apply new training; and they have feedback mechanisms that enable them to reinforce positive learning and correct mistakes	• Outcome 3 ('Improved teacher pedagogical knowledge) to Intermediate Impact ('Improved teacher effectiveness in classroom')
2d	• Teachers have access to sufficient **materials**, their **class sizes** are not too large, and the **social and political context** does not prevent the adoption of new teaching practices	• Outcome 2 ('Improved teacher subject content knowledge) and Outcome 3 ('Improved teacher pedagogical knowledge) to Intermediate Impact ('Improved teacher effectiveness in classroom')
3a	Head teachers are either **motivated or incentivised** to identify, incentivise and influence the positive teaching and management practices that TDP promotes - **SUBEB and LGEA inspectors incentivise** head teachers to identify, incentivise and influence the positive teaching and management practices that TDP promotes	• Output 2 ('Training and support') to Outcome 1 ('Improved head teacher SLM') • Outcome 1 ('Improved head teacher SLM') to Intermediate Impact ('Improved teacher

#	Assumption	Relevant links in TOC
	- **SBMCs** and other community actors **incentivise** head teachers to identify, incentivise and influence the positive teaching and management practices that TDP promotes	effectiveness in classroom') via Outcomes 2 and 3
	- Head teachers are **motivated** to identify, incentivise and influence the positive teaching and management practices that TDP promotes	
3b	Head teachers **have the ability to** identify, incentivise and influence the positive teaching and management practices that TDP promotes.	
	- Head teachers are **able to identify and differentiate positive and negative teaching practices** and to provide feedback to teachers to influence the positive teaching practices that TDP promotes	
	- Head teachers are **formally empowered by LGEA/SUBEB to use 'carrots' and 'sticks' effectively** to incentivise and influence the positive teaching practices that TDP promotes	
	- Head teachers are able to find **informal** (i.e. non-government mandated) **ways of using 'carrots' and 'sticks' effectively** to incentivise and influence the positive teaching and management practices that TDP promotes	
3c	Head teachers are able to exercise effective **management of school infrastructure and resources** to facilitate the adoption of the positive teaching practices that TDP promotes.	• Outcome 1 ('Improved head teacher SLM') to Intermediate Impact ('Improved teacher effectiveness in classroom')
	- Head teachers have access to adequate support to **maintain and repair the school infrastructure and resources** that facilitate the adoption of the positive teaching practices that TDP promotes	
	- Head teachers **manage and allocate school resources efficiently** to facilitate the adoption of the positive teaching practices that TDP promotes	
4	**Cluster meetings** are conducted regularly and effectively, and teachers are motivated to participate in them. Learning from cluster meetings is further reinforced by **peer-to-peer learning** among teachers in schools.	• Quality of TDP in-service activities and Outputs 2 ('Training and support'), 3 ('Materials development'), 4 ('Technology use and management'), and 5 ('INSET programme implementation')
	Both printed and audio-visual training **materials** are fit-for-purpose and appropriate for the levels of skills and knowledge teachers have and the day-to-day challenges they encounter in classrooms.	
	TFs, who constitute the central frontline training workforce for the programme, are appropriately selected, trained, and motivated to deliver their role effectively.	

4.7 Possible limitations of the qualitative research component

4.7.1 Qualitative sampling and generalisability

The qualitative research component is based on a relatively small sample of schools. As explained in the evaluation framework (EDOREN, 2014), it will be a major challenge to generate results that have wider application beyond the schools that we visit. The need to specify a sample in advance means that we cannot guarantee that all possibilities will have been exhausted and all processes understood by the end of the study. Rather, important questions are likely to remain, which can be addressed by subsequent rounds of quantitative and qualitative research.

Sampling for this component is purposive: its aim is to include schools with particular characteristics, rather than being representative of all schools in the three states. Representativeness was further compounded by the inability of the research teams to access all or any LGA as they wish, due to security concerns. Thus, more remote schools farther away from the state capital were less likely to ultimately feature in the final sample of schools. Furthermore, the evaluation framework suggested qualitative research in control schools. However, this was not in fact possible because of resource constraints and a desire to focus on the treatment schools.

The qualitative research is not designed to produce results that are generalisable in the same sense as quantitative data. Generalisability derives from linking qualitative findings to the TOC and to findings from the quantitative research. In some cases it may be more appropriate to talk of whether findings are transferable rather than generalisable: we need to investigate context in order to construct an argument that a finding in one setting is likely to apply in another. The risk of visiting atypical schools and gaining an incorrect or incomplete understanding of the relevant processes remains, but is mitigated by visiting several schools in different states and by paying close attention to ways in which the context of each school may be atypical.

4.7.2 Structured and unstructured research instruments

Qualitative research uses instruments (interviews or discussion guides, observational tools, etc.) which are generally less structured than those used in quantitative research. They leave scope for the interviewer and respondent to shape the research. For example, the interviewer can ask further questions that occur to him or her, in response to an interesting or unexpected response from the research participants. This can help capture impacts or explanations that were not anticipated, but makes qualitative research hard to reproduce and subject to researcher bias. We managed this limitation by using a mixture of relatively structured methods (e.g. semi-structured interviews) and less structured methods (e.g. interviews about teachers' photo diaries, where the teacher can guide the discussion). While the dialogue may be unstructured, the researchers applied structured methods in recording and analysing the discussion: for example, through application of structured templates organised by thematic categories for note-taking, and use of the evaluation matrix (Annex A.4) to provide a framework for analysing the research. A reflective approach, with a mixed team of international and local researchers, and discussion about findings at the end of each day was intended to reduce bias from individual researchers. However, qualitative research inevitably involves greater implication of the researchers' (and participants') own perspectives. This more embodied, personal approach compared to quantitative research should be seen as a strength as well as a limitation.

4.7.3 Sensitive issues

Some issues may have been sensitive for our participants to discuss in the schools. For example, head teachers and teachers were likely to be nervous at first about revealing potentially negative, critical, or self-damaging views or information. A longer-term engagement with the participants would have helped us to gain their trust, but this was not possible given our resource constraints and the need to visit a range of schools in the three states within a limited period of time. Nevertheless, continued presence in the school for three days gave researchers time to gain some trust from the participants, and also allowed time for informal talk and observation as well as more structured discussions. The answers to some questions posed to teachers may have involved overt or tacit criticism of the head teacher or other teachers, if they were frank, and it has to be acknowledged therefore that researchers may not have received comprehensive and honest answers on these questions. Wherever the environment allowed, researchers ensured that discussions took place in private so that only the participants in each part of the research were present. Researchers also used strict codes of data confidentiality and reassured participants that their responses would not be shared more widely. Although the qualitative research remains limited in its ability to explore sensitive issues, it is likely to be stronger in this respect than quantitative research, where there is very little time for researchers to gain the trust of the respondents or to probe evasive or incomplete answers.

4.7.4 Language issues

Interviews were conducted in Hausa and both interviewers and note-takers were Hausa speakers. Important points were interpreted for the benefit of non-Hausa-speaking international researchers, to enable them to guide parts of the discussion. The research team performed preliminary analysis of the findings each day based on their notes and recollection, but the main analysis was conducted by the international researchers upon their return to the UK. The conversations were recorded and the transcripts translated for full analysis of the data. However, there is some risk in this process of inaccurate or incomplete translation. The inclusion in the team of a majority of researchers who were fluent in both Hausa and English was essential in managing this risk. Hausa-speaking researchers checked transcripts in English and Hausa to ensure accurate translation. Researchers were mindful of the need for precision in interpretation in the field, and carried out discussion amongst each other to ensure a clear shared understanding. An added complication was variation and dialects in the Hausa language found in different states (particularly Zamfara). In order to mitigate this risk, to the extent possible, national researchers were selected on the basis of having extensive experience of working in the study areas.

5 Permits, consent, confidentiality and datasets

Conducting fieldwork requires high ethical standards, to ensure that expectations are not raised, confidentiality is maintained and respondents are never forced to participate or encouraged to speak about subjects that may be traumatising (especially for children). Both quantitative and qualitative data collection research proposals for this impact evaluation were passed through both OPM's ethical review board and the National Research Ethics Committee (NREC) Abuja.

The quantitative fieldwork was carried out by field teams made up of national enumerators and field supervisors, supported by staff from the OPM Nigeria office. The interviews with head teachers and teachers and pupil testing were conducted in Hausa. After discussions with the TDP state staff the OPM Nigeria office team arranged for the delivery of letters of permission to visit schools to the SUBEB officials and Education Secretaries concerned in sampled LGAs. Sending the permit letters was not considered sufficient to ensure the Education Secretaries had read and agreed to the school visits. Therefore, follow-up phone calls were carried out to confirm that they had received the letters seeking permission to visit schools from the SUBEBs, and that they understood the purpose of the research and allowed the field teams to visit schools in their LGA.

Informed written consent was sought from all participants for the quantitative research. Given that the baseline surveys were school-based (and not home-based) it was not possible to seek consent from pupils' parents, and hence consent from the head teachers (as the 'guardians' of the pupils whilst they are in school) and from the pupils themselves were sought. Verbal assent was sought from children, and the head teacher signed a written consent on their behalf. When they arrived at schools, the team supervisors started by introducing themselves and their teams to the head teacher, explaining the purpose of the visit and the time that would be required to complete the survey. The enumerators introduced the study and interviews/texts to the head teacher and to all the respondents (pupils and teachers) and were given the option to refuse to participate in the study. If a respondent was reluctant and/ or further explanation was requested, the enumerators were trained to be as exhaustive as possible in explaining the study and its purpose. No head teacher or pupil declined to participate in the survey.

The qualitative fieldwork was carried out by a team made up of national researchers and international staff from EDOREN/OPM. KIIs and FGDs were conducted in Hausa. The field teams undertook all possible measures to keep disruptions of the school day to a minimum by ensuring that head teachers were informed in advance of the dates of the school visits and regarding which types of FGDs and KIIs would take place. The FGDs and KIIs were recorded after the informed written consent of participants was granted. The sequencing of KIIs and FGDs was also – as far as possible – organised in cooperation with school members, in order to minimise disruption to school life and to ensure smooth running of the research. KIIs and FGDs frequently took place outside the school building in order to minimise disruption within teaching spaces.

Informed written consent was sought from all participants at the state, LGA and school levels for the qualitative research. The aims of the research and their ability to withdraw consent at any point during the interviews or discussions were explained to participants. In order to ensure that participants were comfortable with the procedure researchers would read out the explanation and ask participants whether the information provided was clear. Participants were invited to either end or temporally interrupt the interview or discussion if additional questions or concerns arose.

The fieldwork included FGDs with children. The children participating in the research were boys and girls from Grade 3. There is some debate in the development community regarding who is in a

position to provide consent for research conducted with participants who are young children. For ethical reasons, the team decided to gain both the consent of head teachers and children. As a first step, head teachers were asked to provide written consent that they were willing to allow the children to participate in the research. If permission was granted, the children were asked to also provide their assent to participating in the research. At both stages the nature of the research was explained and it was made clear that children were under no obligation to participate.

The evaluation upholds several aspects of DFID's human rights approach (especially participation and inclusion) (Piron and Watkins, 2004) through rigorous training on, and practice of, ethical standards during data collection. This includes seeking consent from respondents, facilitating participation of respondents irrespective of disability status, training gender-balanced data collection teams, among other considerations.

Though not totally avoidable, the interviews were scheduled by data collectors in order to minimise any interruption to the normal flow of activities in the school and the need for teachers and pupils to stay beyond school hours. No monetary incentives were given to respondents for participation in the study. Each school that was part of the qualitative study received a gift (an inflatable globe) and a box of chalk as a token of thanks for their time and participation. The total value of these gifts was well under £3. Participants in the FGDs and KIIs for the qualitative fieldwork received refreshments and, additionally, children received a pencil, two to three crayons and an eraser. Children who participated in the qualitative survey also received similar items and refreshments. The expected perverse effect of these gifts on respondents is considered to be minimal.

This independent impact evaluation is being carried out by EDOREN and is intended for primary consumption by TDP and DFID Nigeria. As such, the final ownership and copyright of the data, analysis and reports rests with EDOREN, which is managed by OPM. However, all outputs (especially reports) produced under this evaluation – by joint agreement – will be co-branded to bear EDOREN, UK Aid and TDP logos.

Data ownership is defined by DFID's contracts with OPM for EDOREN, and with Mott McDonald for TDP. It is EDOREN's understanding that the data collected are co-owned by Mott McDonald, OPM, and DFID. As stated in the TDP evaluation framework, the clean, anonymised evaluation datasets and metadata will be made publically available, probably on the EDOREN website and in the World Bank micro-databank (subject to DFID approval), so that researchers can replicate and extend the evaluation analysis, in line with DFID's Open Access policy .

Intellectual property rights in respect of any materials produced by EDOREN (such as evaluation reports, policy briefs etc.) are the property of OPM. However, OPM has granted DFID a worldwide, non-exclusive, irrevocable, royalty-free licence to use all this data and material.

All personal data collected as part of this survey are available only to authorised individuals for analytical purposes and are handled using data protection best practices. Each respondent has been assigned a unique identifier that is used to analyse the data. All cleaned and documented datasets, anonymised by removing personal information that could be used to identify respondents, related to the baseline study will be made public through the EDOREN website and World Bank micro-databank (subject to DFID approval) to enable national researchers, research students and other education stakeholders to access and use the impact evaluation data to conduct additional analysis and research. All data have been backed up and are stored in an 'OPM

Stats archive'. OPM will store all original data and transcripts for three years, after which time they will be destroyed.

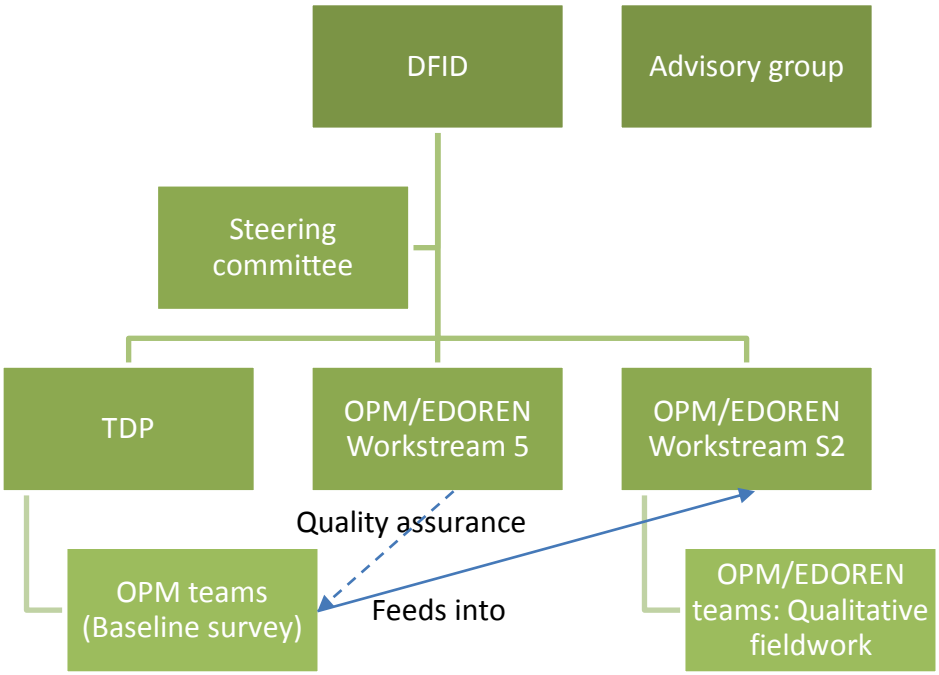
6 Evaluation governance, management, independence and ethics

The evaluation of the TDP is designed and implemented by EDOREN, a DFID-funded project, and therefore has clear accountability to DFID through project reporting. Thus far the evaluation has been, and will continue to be, governed by a tripartite steering group composed of DFID staff (the relevant education adviser and results adviser), a TDP representative (the TDP research and evidence output lead), and EDOREN representatives (the EDOREN Country Director and TDP workstream lead). The independence of the evaluation is assured through transparent and rigorous peer review.

6.1 Governance and management

This evaluation is managed and implemented by EDOREN, a DFID-funded project managed by OPM, and, specifically, EDOREN's Workstream 2: TDP evaluation and support activities. EDOREN Workstream 2 has been responsible for the design of the evaluation framework and the implementation of the evaluation. The evaluation drew on the quantitative data collected by teams from OPM Oxford and Nigeria under contract (for the baseline survey) directly from Mott MacDonald, which is the implementing agency for TDP, as well as on qualitative data collected by staff from OPM Oxford and local researchers contracted by EDOREN under the auspices of EDOREN's contract with DFID. The baseline survey group has reported regularly to the EDOREN Workstream 2 team to ensure that the baseline meets the requirements of the evaluation framework, as well as of the TDP. In addition, EDOREN's Workstream 5 on policy impact and statistical quality independently reviewed the quality of quantitative data collections conducted by OPM teams.

Figure 4 TDP evaluation governance arrangements



The overall evaluation, including follow-up surveys, will be ultimately governed by DFID. Specifically, this entails that the EDOREN Workstream 2 team reports to the DFID Nigeria education team on progress towards evaluation objectives. This will take place through regular EDOREN quarterly written reporting to DFID, and six-weekly verbal project management updates.

This process means that the EDOREN TDP evaluation team provide fortnightly updates to EDOREN management (the project manager and Country Director), and EDOREN management can therefore provide additional updates to DFID as required.

In addition, as per the recommendations of the TDP evaluation framework (EDOREN, 2014), the evaluation steering committee, composed of DFID Nigeria education and results advisers, TDP and EDOREN staff, has been meeting regularly to discuss evaluation progress, and is responsible for peer review and quality assurance (in addition to EDOREN's internal quality assurance processes). The designated EDOREN Workstream 2 lead provides updates on the evaluation and reports to this committee every six months. An advisory group of secondary users is also expected to be constituted. This would include the groups above, plus the state ministries of education, state Commissioners of Education, SUBEBS in the six TDP states, the Federal Ministry of Education, UBEC, the NERDC, and ESSPIN and GEP3. This advisory group would meet annually to discuss evaluation progress as reported by the EDOREN Workstream 2 lead for the TDP evaluation, and would offer advice to next steps.

6.2 Independence, credibility and usefulness

This section briefly discusses how the evaluation achieves the principles of independence, credibility and usefulness, and appropriate participation.

In this case, the principles of independence are difficult to achieve, because EDOREN was providing advice to the TDP, particularly on results and evidence, but also to some extent (through the evaluation framework) on the nature of its interventions. EDOREN was also both responsible for designing the evaluation framework and for conducting the evaluation. While this is for good reasons (EDOREN's remit is to ensure that data and research are used as much as possible through DFID's education portfolio in Nigeria), these EDOREN activities raise potential concerns about the impartiality and independence of the TDP evaluation. EDOREN's approach to mitigating these concerns is to ensure that: first, all evaluation outputs are externally peer reviewed; and second, all evaluation frameworks and reports are approved by SEQAS; third, data and analytical approaches for both quantitative and qualitative data are transparent (i.e. publicly available for scrutiny); and fourth, to contract independent annual review teams as per EDOREN's inception report.

To improve credibility, EDOREN will ensure that both evaluation processes (e.g. datasets, analysis approaches) and outputs will be publicly available for external scrutiny, with the required confidentiality safeguards.

The evaluation is designed to be useful. As the main purpose of the evaluation is to learn about what works for improving teacher effectiveness and learning in low resource contexts, the evaluation outputs in general do not require a specific timeline. However, the baseline and follow-up rounds of survey will provide important evidence for adjustments to TDP Phase 2. EDOREN's approach to ensuring the relevance of the evaluation questions and proposed priorities is to consult widely before the finalisation of evaluation outputs, including the evaluation framework, baseline reports and future outputs. Credibility and reliability will be achieved through the steps outlined above, and through the use of high quality expertise in conducting the evaluation.

In addition, TDP's evaluation will aim to fulfil the principles of usefulness and participation set out above through strategic and sustained stakeholder engagement and information dissemination at various stages of the evaluation process. This is also part of the TDP's Research and Evidence

(R&E) approach (McCormick, 2014), and the evaluation's proposed approach to stakeholder engagement, for which see Section 7. Thus far, the baseline survey results have been presented to an international audience at the UK Education and Development Conference (also known as UKFIET) at Oxford in September 2015, as part of the panel on teachers in northern Nigeria, and at a one-day workshop in October 2015 at DFID's offices in Abuja to an audience comprising TDP Abuja and state staff, TDP annual reviewers, DFID education, governance and results advisers, ESSPIN, GEP, and EDOREN Workstream 2 researchers.

In addition, ethical evaluation principles are important, and run through the evaluation design. These were already discussed in relation to consent and confidentiality in Section 5. One area of particular importance for this evaluation has been ensuring that children are asked age-appropriate questions and that appropriate consent is sought. In this regard the evaluation has followed UNICEF guidelines (UNICEF, 2013) on conducting research with children. In addition, both quantitative and qualitative data collection research proposals have passed through both OPM's ethical review board and the NREC in Abuja.

6.3 Addressing equity, poverty and exclusion

The intervention design does not reflect considerations of equity, poverty and the exclusion of specific groups as such. Hence, samples in the quantitative survey were drawn randomly to reflect the gender, age, ethnicity etc. characteristics, at least, of the areas they were meant to be representative of; and data collection teams (which were themselves gender-balanced as much as possible) were trained to demonstrate sensitivity to diverse groups. This includes children with disability, who if randomly selected, were included in the sample irrespective of their disability status even if there are pressures from teachers and head teachers to replace a disabled pupils. At the stage of analysis and reporting, essential summary statistics were then reported by relevant sub-groups, e.g. by gender and socio-economic status.

The evaluation upholds several aspects of DFID's human rights approach (especially participation and inclusion) through rigorous training on, and practice of, ethical standards during data collection. This includes seeking consent from respondents, facilitating participation of respondents irrespective of disability status, and training gender-balanced data collection teams, among other considerations. Sections 3 and 4 above discussed the process of designing and piloting the instruments at multiple sites to ensure the appropriateness of the final instruments for the respondents in questions.

7 Stakeholder engagement and dissemination of evaluation results

In this annex we discuss the strategy for engaging stakeholders throughout the various stages of the evaluation process, and plans for disseminating the results of the evaluation. In evaluating development interventions it is important to engage beneficiaries and stakeholders at all stages of the evaluation, including the design of the framework, the development of tools and data collection. The first and second sub-sections of this section outline the roles of stakeholders, and their involvement in the development of the evaluation framework. In the final sub-section we discuss the plan for communicating evaluation results and key considerations.

The main objectives of the stakeholder engagement and communication strategy are to ensure ownership by stakeholders throughout each stage of the process, to facilitate learning and sustainability of results, and to ensure the effective use of evidence generated by stakeholders.

7.1 Role and involvement of stakeholders

Understanding the political, organisational and technical context is necessary when evaluating development interventions. Understanding the needs and issues of beneficiaries and stakeholders is important at all stages of the evaluation process, from design of the evaluation framework to the subsequent implementation and, finally, the communication of evaluation findings. Involving stakeholders at each stage of the evaluation will help to ensure that: first, there is a common understanding of the problem being addressed by the intervention; second, the right questions are asked; third, the questions are appropriately phrased; and, finally, the methods are agreed and understood. TDP's evaluation will use an interactive and consultative participatory methodology to engage stakeholders at various stages of the evaluation process.

The main purpose of this evaluation is to identify the causal pathways through which the provision of support to teachers in resource-poor contexts improves their effectiveness and, ultimately, student outcomes, i.e. learning what works to support teachers in resource-poor contexts. Therefore, the principal stakeholders for the evaluation are: DFID Nigeria, TDP, the state governments of the Katsina, Jigawa and Zamfara – especially state Ministries of Education and SUBEBs; NCCE and CoEs; other education projects, especially GEP3 and ESSPIN; organisations (primarily DFID) seeking to improve teacher competence and subsequently learning elsewhere in the world, international researchers on education, and Nigerian education policy-makers and researchers.

Involving key stakeholders in the design of the evaluation framework and the conducting of the evaluation provided perspectives that helped to ensure a credible, high quality and useful evaluation, and contributed to the programme logic and the formulation of key evaluation questions.

It is important that all stakeholders concerned feel involved in the evaluation and are concerned with the results generated. Consulting stakeholders at the early stages thus ensured greater ownership of the evaluation by increasing knowledge about and support for the evaluation. This in turn results in higher quality of data collection and research, and utilisation of the evaluation findings.

Key stakeholders were / will be involved in the conducting of the evaluation in the following ways:

- Interactive discussions with TDP to finalise data collection tools and methodologies to ensure all necessary questions were asked and there was consensus on every aspect of the approach.
- Feedback of results to TDP and DFID Nigeria to ensure they are taken into account for other (current or future) teacher development programmes aimed at improving teacher competency and broader learning outcomes.
- In addition, the involvement of parents and teachers in the evaluation, and particularly the dissemination, of findings could have beneficial impacts on the sustainability of the programme. The evaluation team will discuss further with the TDP's R&E component the possibility of joint dissemination activities in the TDP implementation states, as it is likely to be outside EDOREN's capacity to engage directly with these stakeholders.

  Feedback of results to DFID Nigeria, international organisations (primarily DFID) seeking to improve teacher competencies elsewhere in the world, and international education policy-makers and researchers, to help make more informed decisions about whether to apply the TDP model of teacher development in other contexts. As has been said, thus far the baseline survey results have been presented to an international audience at the UK Education and Development Conference (also known as UKFIET) at Oxford in September 2015, as part of the panel on teachers in northern Nigeria, and at a one-day workshop in October 2015 at DFID's offices in Abuja, to an audience comprising TDP Abuja and state staff, TDP annual reviewers, DFID education and governance.

Stakeholders were/will be involved in two broad phases: development of the evaluation framework (2014) and communication of the evaluation findings (this process started in late 2015).

7.2 Stakeholder involvement in framework development

Stakeholders were involved in the development of the evaluation framework in various ways:

- DFID Nigeria. Consultative discussions were held with the DFID team to understand the main aims of the intervention and the key activities within each component, and to understand the main uses of an evaluation. Discussions were also held with DFID to clarify the main uses and purposes of the evaluation. Education advisers and the results adviser had final signoff on the evaluation framework, before forwarding to SEQAS for quality assurance.
- TDP R&E Lead Bukola Oyinloye was consulted on the development of the framework.
- International researchers within OPM and EDOREN's networks were consulted on the development of the framework.

7.3 Strategies to communicate evaluation findings

Effective communication of evaluation findings will ensure that the results reach the concerned stakeholders and are actively taken into consideration by them. The strategy for the communication of evaluation findings has been developed in line with TDP's communication strategy and EDOREN's overall dissemination plan.

EDOREN's strategy will keep in mind the uses of evaluation and technical abilities of users in determining the best ways of communicating information to each stakeholder. It is crucial that the evidence generated is presented in appropriate formats and is available in the spaces and places where the relevant stakeholders are likely to seek out the evidence. The dissemination strategy will therefore include the following steps:

- Ensuring that communication of evidence produced in partnership with other DFID education portfolio programmes is discussed and agreed between the concerned programmes, including determining if any sections may not be suitable for sharing beyond specific audiences.

- Evidence will be made available in formats and styles appropriate to each of the priority stakeholder groups: this is likely to include policy briefs, summary reports, presentations, data visualisations, radio programmes, videos and blogs. Our reports and recommendations will summarise the key findings in non-technical language, supported by technical annexes.

- Ensuring evidence is available and present in the spaces and places that our stakeholders frequent: spaces could be either physical or virtual, places where information can be found whether or not it is being actively searched for. For virtual spaces this includes improving access to key international and Nigerian websites. Conferences and stakeholder meetings will also be organised to facilitate discussions with key stakeholders who might not have access to virtual platforms. Other channels, such as radio, newsletters and local and international media, will also be explored.

- Appropriate strategies of stakeholder engagement will be applied to promote engagement with the evidence and subsequent use: meetings with stakeholders are a key mechanism for sharing knowledge and learning, and for facilitating stakeholder understanding of the evidence and how this might affect their practice or policy-making. Engaging with local and national media to raise their awareness of, and interest in, basic education is also key, as these channels can also influence public views and understanding of education-related policy issues.

Bibliography

Bangladesh DPE. 2013. "National Student Assessment for Grades 3 and 5." Bangladesh Directorate of Primary Education, Ministry of Primary and Mass Education.

Beavis, Adrian, and Rachel Outhred. 2014. "Pre-Literacy Literature Review for EDOREN, TDP and GEP Evaluations." Oxford Policy Management (OPM).

Bennell, Paul, and Kwame Akyeampong. 2007. "Teacher Motivation in Sub-Saharan Africa and South Asia." Department for International Development: Educational Papers. DFID.

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

———. 2015b. "Teacher Motivation: Methods and Results from the TDP in-Service Baseline Survey." EDOREN (unpublished note).

Carvalho, Soniya, and Howard White. 1997. *Combining the Quantitative and Qualitative Approaches to Poverty Measurement and Analysis: The Practice and the Potential*. World Bank Publications.

Cohen, Jacob. 1988. *Statistical Power Analysis for the Behavioral Sciences*. L. Erlbaum Associates.

Deaton, Angus S. 2009. "Instruments of Development: Randomization in the Tropics, and the Search for the Elusive Keys to Economic Development." Working Paper 14690. National Bureau of Economic Research. http://www.nber.org/papers/w14690.

De, Sourovi, Gunilla Pettersson, Peter-Sam Hill, and Robert Morris. 2015a. "TDP Katsina Report." Oxford Policy Management (OPM).

———. 2015b. "TDP Zamfara Report." Oxford Policy Management (OPM).

De, Sourovi, Gunilla Pettersson, and Robert Morris. 2015. "TDP Jigawa Report." Oxford Policy Management (OPM).

Donovan, Jenny, Melissa Lennon, Gayl O'Connor, and Noni Morrissey. 2008. "National Assessment Program–Science Literacy Year 6 Report, 2006." *Ministerial Council on Education, Employment, Training and Youth Affairs (NJ1)*. http://eric.ed.gov/?id=ED534659.

Duflo, Esther, and Michael Kremer. 2004. "Use of Randomization in the Evaluation of Development Effectiveness." MIT - Massachusetts Institute of Technology. http://economics.mit.edu/files/765.

EDOREN. 2014. "TDP Evaluation Framework and Plan." Oxford, UK: Education Data, Research & Evaluation in Nigeria (EDOREN).

ESSPIN. 2013. "Overall Findings and Technical Report of ESSPIN Composite Survey 1 (2012)." ESSPIN 060. ESSPIN , UKAID. http://www.camb-ed.com/Portals/0/Documents/ESSPIN-Composite-Survey-Technical-Report.pdf.

———. 2014. "Education, Conflict and Violence Research: ESSPIN Report 428." Abuja.

Federal Republic of Nigeria. 2004. "National Policy on Education 2004." NERDC Press.

Garbarino, Sabine, and Jeremy Holland. 2009. "Quantitative and Qualitative Methods in Impact Evaluation and Measuring Results." Monograph. March. http://www.gsdrc.org/docs/open/EIRS4.pdf.

Goldstein, Harvey. 2015. "Rasch Measurement: A Response to Payanides, Robinson and Tymms." *British Educational Research Journal* 41 (1): 176–79. doi:10.1002/berj.3170.

Guest, Greg, Kathleen M. MacQueen, and Emily E. Namey. 2011. *Applied Thematic Analysis*. Sage.

Johnson, D. 2008. "An Assessment of the Development Needs of Teachers in Nigeria - Kwara State Case Study."

Johnson, David, and Sergij Gabrscek. 2008. "An Assessment of the Development Needs of Teachers in Nigeria- Kwara State Case Study." ESSPIN/UKAID.

Johnson, David, and Jenny Hsieh. 2014. "An Assessment of the Development Needs of Teachers in Nigeria - Katsina State Case Study."

McCormick, Robert. 2014. "TDP Results and Evidence Strategy." Cambridge Education.

McKenzie, David. 2015. "Tools of the Trade: A Joint Test of Orthogonality When Testing for Balance." *Impact Evaluations*. http://blogs.worldbank.org/impactevaluations/tools-trade-joint-test-orthogonality-when-testing-balance.

NBS. 2012. "The Nigeria Poverty Profile 2010 Report - Press Briefing by the Statistician-General, Monday 13th February 2012." Abuja.

Neves de Jesus, Saul, and Willy Lens. 2005. "An Integrated Model for the Study of Teacher Motivation." *Applied Psychology* 54 (1): 119–34. doi:10.1111/j.1464-0597.2005.00199.x.

OPHI. 2014. "OPHI Country Briefing 2014: Nigeria." Oxford.

Outhred, Rachel. 2015. "Concept Note on Learning Outcome Measurement across EDOREN Evaluations." EDOREN (unpublished note).

Panayides, Panayiotis, Colin Robinson, and Peter Tymms. 2010. "The Assessment Revolution That Has Passed England by: Rasch Measurement." *British Educational Research Journal* 36 (4): 611–26. doi:10.1080/01411920903018182.

Pettersson, Gunilla, and Georgina Rawle. 2015. "EQUIP Tanzania Impact Evaluation Briefing Note 1: Teachers' Knowledge, Behaviour and Support in Some of the Most Disadvantaged Districts in Tanzania." Oxford Policy Management (OPM).

Piron, Laure-Helene, and Francis Watkins. 2004. "DFID Human Rights Review - a Review of How DFID Has Integrated Human Rights into Its Work." ODI.

Sahara Reporters. 2014. "Boko Haram Takes Over Part Of Jigawa State, Shuts Down 30 Schools." *News Rescue*. http://newsrescue.com/boko-haram-takes-part-jigawa-state-shuts-30-schools/#axzz3Udxitbcb.

Siraj, Iram, Brenda Taggart, Edward Melhuish, Pam Sammons, and Kathy Sylva. 2014. "Exploring Effective Pedagogy in Primary Schools: Evidence from Research." https://research.pearson.com/content/plc/prkc/uk/open-ideas/en/articles/explore-eppse/_jcr_content/par/articledownloadcompo/file.res/Exploring%20Effective%20Pedagogy%20in%20Primary%20Schools.pdf.

Taft, Patricia, and Nate Haken. 2015. *Violence in Nigeria: Patterns and Trends*. Cham, Switzerland: Springer.

UNESCO. 2014. *Teaching and Learning: Achieving Quality for All*. http://unesco.nl/sites/default/files/dossier/gmr_2013-4.pdf?download=1.

UNICEF. 2013. *Ethical Research Involving Children*. http://www.unicef-irc.org/publications/706.

USAID. 2013a. "Results of the 2013 Early Grade Reading and Early Grade Mathematics Assessments (EGRA & EGMA) in Bauchi State." Northern Education Initiative (NEI), USAID.

———. 2013b. "Results of the 2013 Early Grade Reading and Early Grade Mathematics Assessments (EGRA & EGMA) in Sokoto State." Northern Education Initiative (NEI), USAID.

Watts, Michael, and Terry Allsop,. 2015. "EDOREN Teacher Management Report (Kaduna and Katsina)." EDOREN.

Westbrook, J. 2013. "Pedagogy, Curriculum, Teaching Practices and Teacher Education in Developing Countries." Department for International Development (DFID). https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/305154/Pedagogy-curriculum-teaching-practices-education.pdf.

Annex A Terms of Reference

The TORs for this impact evaluation have developed over several stages. Initially a TOR was developed jointly by TDP and EDOREN to design the evaluation framework (EF) for the impact evaluation of TDP's in-service teacher training component (Annex A.1), based on which EDOREN developed the EF that laid out the overarching evaluation questions to be investigated by the impact evaluation (Annex A.2). Guided by these overarching questions, an analysis plan was developed by EDOREN for the quantitative baseline survey which defined key quantitative indicators to be reported on (Annex A.3). Further a concept note was developed by EDOREN to outline the qualitative baseline survey's evaluation matrix (Annex A.4).

These are all outlined below and EF will be submitted as a supporting document to SEQAS for review.

A.1 TORs for evaluation design

A.1.1 Summary

Education Data, Research and Evaluation in Nigeria (EDOREN) generates new evidence and understanding of how best to support equitable access and improved learning outcomes for all Nigerian children through innovation and sustainable education systems development. EDOREN seeks a team of short-term consultants to design a robust evaluation framework and plan for the Teacher Development Programme (TDP). The detailed tasks in these ToR will be refined during the mission on the basis of discussions with the relevant stakeholders. The team will interact closely with the EDOREN Project Manager, DFID, the TDP team, and other stakeholders in education in Nigeria and TDP states.

A.1.2 Background

EDOREN embeds high quality data, research and evaluation in DFID Nigeria's education portfolio and in the education policy of partner Nigerian States through:

- The provision of complex and long-term education research, statistical support and political economy analysis
- Building national capacities and incentive to generate and use data
- The provision of better quality information for policymakers

Workstream 1 will lead to a demonstrable improvement in the quality of DFID basic education project evaluation through conducting and providing support to annual and in-depth reviews and evaluation of three DFID Nigeria education projects: The Girls Education Project 3 (GEP3), DEEPEN and the Teacher Development Programme (TDP).

Effective project review and evaluation is critical to accountability, project improvement and learning for the future. Demonstrable improvement in the quality of DFID basic education project evaluation will contribute to stronger accountability of projects to DFID, to strengthening the design of current projects, and to influencing the development of DFID's education strategy post 2016.

EDOREN will perform strong external review and evaluation functions for GEP3, DEEPEN and TDP. These will follow DFID evaluation standards and processes, which mostly include the use of OECD DAC (Organisation for Economic Cooperation and Development, Development Assistance Committee) criteria for rigorous impact evaluations.

The Teacher Development Programme (TDP) is a 6 year (2013-2019) project funded by UKaid/DFID that will support Federal and State institutions which are responsible for the pre-service and in-service development of teachers in six states in Northern Nigeria. It will have the opportunity to make permanent improvements to the quality of teaching in Nigeria, affecting the life chances of millions of young Nigerians. Through a technology-enhanced, innovative teacher training model, the TDP will support governments to design and implement strategies to improve the standard of in-service and pre-service training and support the continuing professional development of teachers in Nigeria.

The purpose of the TDP is to improve the quality of teaching and learning of children in basic education. The main objective is to improve the skills of teachers in the three core curriculum subjects of English, maths, science & technology and produce better teachers through a combination of pre-service and in-service interventions while the specific objectives are to:

- Reform the process of pre-service production of teachers;
- Establish a sustainable system of in-service development of teachers;
- Develop innovative and multifaceted ways of identifying and then addressing constraints and help teachers achieve greater job satisfaction through enhanced classroom performance; and
- Improve the evidence base about what works and what does not in both pre-service and in-service teacher development.

The TDP Results & Evidence strategy (McCormick 2013), on which EDOREN has provided some initial advice, has three main areas of work: Programme evaluation; Programme monitoring, of in-service and pre-service activities; and other studies that will be conducted to bolster the evidence base on education in Nigeria.

For Programme evaluation, the Programme will measure progress and will report against the Logframe in terms of impact, outcomes and outputs. To do this, the Programme has identified different indicators of success at the impact, outcome and output levels. At the impact level, the TDP's aim to improve student learning in target schools will be measured by student learning outcomes, school survival rate, as well as students' net attendance rates.

At the outcome level, Outcome 1 will be measured by 1) change in teachers' use of positive interaction; 2) teachers surveyed reporting improved motivation; 3) the change in teacher absenteeism; and 4) change in teacher subject and pedagogic knowledge competency. Outcome 2 will be measured by 1) colleges' performance in the key areas of the QA assessment; 2) the extent of lecturers' use of activity-based, learner-centred approaches; and 3) the increase in the level of student teacher satisfaction with the support received during teaching practice.

At the output level, Output 1 will report the number of teachers trained; the number of head teacher trained in management and staff development; and the number of teacher facilitators

trained. Output 2 will report the number of CoE staff attending skills updating sessions, the change in the amount support received during teaching practice, and the change in the number of lecturer-student contact hours. Finally, Output 3 will report the number of studies completed; the number published; and the use of its evidence by policy-makers.

The measurement of the above will comprise of the use of various instruments/surveys in in-service (Monitoring of Learning Achievement (MLA) surveys for English, Mathematics and Science & Technology for primary students; TDNA for primary teachers of English, Mathematics and Science and Technology; Classroom Observation; and Teacher Motivation) with that of pre-service (student-teacher survey). Pre-service progress will be further measured through case studies of various intensities while additional qualitative studies will be done to complement the in-service survey.

DFID has recently asked EDOREN to design an evaluation framework for TDP, aligned with the existing M&E strategy. This is covered in these TOR.

EDOREN will also conduct the annual reviews of TDP in 2015 and 2016.

A.1.3 The Objective

The objective is to produce an evaluation framework and plan for TDP, with a draft by the 30th June 2014. This evaluation framework should:

- Be based on the DAC criteria for the evaluation of development assistance, including permitting a rigorous evaluation of TDP's expected impact on learning outcomes;
- Meet DFID prescriptions for an evaluation framework and plan as set out in their evaluation handbook, Business Case, How To note, Evaluation Quality Assurance and DAC ;
- Use the most rigorous evaluation approaches available, as discussed by organisations such as 3ie;
- Build on work already conducted by TDP and EDOREN;
- Build on evaluations of teacher development projects carried out elsewhere (see e.g. http://www.eiabd.com/eia/index.php/2012-10-11-09-41-47/research-publication/research-report/baseline-reports for baseline reports on a similar project in Bangladesh); and
- Be feasible to implement given the resources available to TDP and EDOREN

A.1.4 Recipient

The recipients of the services will include but not be limited to DFID Nigeria, TDP, and the Governments of States in which TDP works, and Nigeria.

A.1.5 Scope of the Task

The evaluation framework and plan will set out a comprehensive approach to the rigorous evaluation of TDP and its key components, particularly in terms of 1) identifying whether TDP is generating its desired impact, outcomes and outputs as set out in the agreed DFID logframe and 2)

establishing causality. The evaluation framework should also enable the assessment of TDP by the DAC criteria: relevance, efficiency, effectiveness, impact and sustainability.

Given the methodological complexity of evaluating TDP as a whole (with work at both the school level and through advocacy), it is expected that the evaluation framework will use a theory-based approach, including mixed methods (White 2009).

However, attention should be given in the framework to the possibility of evaluation designs that will also allow the development of counterfactuals and rigorous impact attribution. The evaluation framework team will need to discuss and agree the feasibility of these designs with the TDP R&E team before proposing them. Following discussions with the TDP R&E team and using the TDP M&E strategy and logframe, the team will evaluate the in-service component using a quasi-experimental design, and the pre-service component using research approaches with smaller 'N's. The feasibility of both of these approaches will need to be assessed.

If the evaluation framework includes surveys, which seems likely, the team must propose solutions to the complex sampling and fieldwork quality issues that surround surveys in northern Nigeria, working with data quality experts from EDOREN (in particular Mary Strode, Matthew Powell and David Megill). David Megill has already worked with the TDP R&E team to develop an initial sample of 56 schools for the TDP in-service baseline evaluation. His sampling note, and the draft TOR for this baseline (see references), would be important parameters for any surveys that are used for the evaluation.

In particular, the surveys will need to solve the difficulties of selecting schools in an experimental fashion, given that in Jigawa TDP schools are selected from the 500 (out of 2,000) where ESSPIN are not working, and in Katsina and Zamfara, TDP schools are selected from the LGAs where GEP3 is not working (this is to be confirmed). The evaluation design, if it includes surveys in all three states to answer evaluation questions, will need to be very clear about what these surveys are representative of. Quasi-experimental approaches could attempt to match schools using data from the Annual Schools Census, but these data may have some quality problems, especially in Katsina and Zamfara.

A further difficulty with surveys is that the TDP will not work with every teacher in a school, but only four. In some schools there are up to 50 primary teachers. It is expected that teachers will be selected by head teachers or local government authorities; this leaves a problem of how teachers in control schools could be selected.

There is an important attempt in Nigeria to harmonise data collection approaches, and there are currently instruments available to measure numeracy and literacy at p2 and p4, Teacher Development Needs Assessments for English and maths at p1-6 and JSS1-3, and approaches to classroom observations. These should be used, or developed where necessary, rather than designing additional indicators.

The evaluation framework must also take note of TDP's existing plans for Monitoring and Evaluation, which are set out in the Results and Evidence strategy. This currently proposes a series of surveys at the school level (TDP 2014). . The evaluation framework should form a view on the most appropriate design and use of this survey, and whether this is the most rigorous way to evaluate TDP given resources available, and what else might need to be added.

Evaluating efficiency and sustainability are notably challenging, and the evaluation framework should propose solutions to these challenges that generate viable information. Evaluating efficiency may require engaging with the TDP project and other stakeholders to ensure that they are able to provide the relevant costing information. This may require an examination of the TDP Value for Money strategy. Alternatively, a process may need to be established. It may also require establishing benchmarks for similar Teacher Development programmes.

Evaluating sustainability may require reliance on a theory-based approach, but may also involve conducting research after the project ends, especially given the objective of improving government decisions around teacher effectiveness and efficiency through research. If the evaluation framework proposes this, it should also propose a practical solution to how this could be done.

The evaluation framework should also give consideration to how the evaluation will be used and communicated, in particular how key stakeholders in teachers in Nigeria will be involved throughout the evaluation process.

The framework and plan will include the following elements:

1. *What is the purpose of the evaluation?*
2. *Who will be the key users of the evaluation?*
3. *When (and how often) should the intervention be evaluated?*
4. *What are the key evaluation questions, organised by DAC criteria and drawing on the logical framework and theory of change?*
5. *What is the evaluation framework? This should turn the questions into measurable indicators and targets, with sources of information for each indicator.*
6. *What design and methods are envisaged? This needs to address the difficulties of counterfactuals, attribution, rigour, sampling, and data availability, as well as the resources available for the evaluation from both EDOREN and TDP.*
7. *Will the baseline data and monitoring strategy provide the data necessary to answer the evaluation questions?*
8. *How does the evaluation approach fit with the existing evidence base in support of the intervention?*
9. *What are the roles of stakeholders and how will they be involved?*
10. *What is the strategy to communicate the evaluation findings?*

The work of the framework team will be split into three phases.

1. Preparatory phase

Objective: get good understanding of 1) TDP programme (interventions, logframe, theory of change, monitoring and reporting framework, value for money strategy), 2) context, 3) existing data, and 4) existing evidence in order to define the objective of the evaluation and identify opportunities and limitations for an evaluation.

Key activities:

1. Draft intervention factsheets (templates will be shared with the team)
2. Review and assess logframe and theory of change
3. Review indicators and reporting needs; draft indicator factsheets
4. Conduct review of contextual factors that may affect 1) the object of the evaluation and 2) the process of the evaluation
5. Draft existing data sources/tools fact sheets (templates will be shared with the team)
6. Review existing evidence base for TDP interventions

2. Conceptual design of core evaluation framework

Objective: develop and agree upon evaluation framework including evaluation questions, approaches/designs, methods and activities.

Key activities:

1. Identify purpose, needs and users/stakeholders of evaluation
2. Specify evaluation questions, evaluation criteria and scope of evaluation
3. Specify evaluation types/approaches/designs
4. Specify evaluation methods and activities, including data collection methods (draft an evaluation matrix)

3. Development of facilitatory evaluation processes

Objective: specify processes and resources that need to be in place to achieve a timely and quality evaluation.

Key activities:

1. Develop evaluation workplan/timeline
2. Establish evaluation management/governance structure/processes
3. Develop ethical and quality evaluation standards
4. Develop communication and reporting plan
5. Define and agree upon evaluation resources

A.1.6 Deliverables

The deliverables for *the TDP evaluation framework* are set out in the table below.

Description of deliverable	Proposed date
Draft evaluation framework including 2 page executive summary and full indicator framework annex	30th June 2014
Draft evaluation framework including 2 page executive summary and full indicator framework annex	One week after receiving DFID comment (expected final framework submission date – 23rd July 2014)

A.1.7 Timeframe

The framework is to be completed by June 30th. The time allocations are as follows:

a. Team Leader – Evaluation Expert – 8 days (Ian MacAuslan)

b. Impact evaluation Experts – 10 days (Alex Hurrell/Tom Pellens)

c. Analysts/Nigeria education experts - 20 days (Sourovi De/Ifeatu Nnodu)

Ian MacAuslan will be responsible for ensuring that the required inputs are made by all consultants in order to produce the final deliverables.

A.1.8 Proposed skills mix of the team

A team leader with expertise in impact evaluation, and particularly in education, will be responsible for oversight of the review. Additional experts will support the development of the framework. The team will include a teacher education expert, and a team of analysts in Nigeria and the UK. Additionally, the team will need to have the following:

- Strong knowledge and awareness of DFID's high level policy role and strategic direction
- Excellent knowledge and extensive experience in Monitoring and Evaluation
- Experience of leading and managing DFID review processes
- Excellent communication and report writing skills

A.1.9 Coordination and Logistics

Coordination with the rest of EDOREN should take place through Ian MacAuslan.

Coordination with the TDP R&E team should take place through Ian MacAuslan, but with some responsibility put on the team when they are in Nigeria to liaise directly with TDP in a way that does not involve over-burdening TDP.

The framework development is partly desk-based and partly Nigeria-based. Desk-based interviews are to be conducted via Skype or telephone.

Documents are available from the EDOREN Google Drive and the Zotero literature database, and team members should have access to this. For any problems, contact Florian.Friedrich@edoren.org

The consultants or their firm will be contracted through the Education, Data, Research and Evaluation in Nigeria (EDOREN) programme, with a contract held by Oxford Policy Management. The EDOREN Abuja office will provide logistical support as necessary in Nigeria, and Oxford Policy Management's Oxford office will provide support internationally.

A.2 Overarching evaluation questions (as proposed by the EF)

OECD DAC Evaluation criteria	Key Question	Formative/summative and timing of answer
Relevance	Does TDP's in-service training approach and design address needs, priorities and constraints of the primary teachers in northern Nigeria?	Formative - 2016
Relevance	Are TDP's assumptions correct globally and particularly for the Nigerian education and policy context?	Formative - 2016
Effectiveness	Has TDP's in-service output led to changes in the effectiveness of teachers in target schools for primary 1-3?	Formative - 2016
Impact	Has TDP caused changes in student learning in English, maths and science & technology in target schools?	Summative - 2018
Efficiency	Does the TDP offer value for money in terms of the cost of impacts, were results achieved on time and to plan, and how does TDP's organisational set up facilitate delivery?	Summative - 2018
Effectiveness	Has TDP led to changes in teacher effectiveness?	Summative - 2018
Sustainability	Are TDP's impacts on teacher effectiveness sustainable without further DFID support?	Summative - 2018

A.3 Definitions of key quantitative indicators

Table 18 provides an overview of the indicators presented in Volume I. Column 1 (EF reference) provides the code used in the Evaluation Framework for each indicator (EDOREN, 2014); column 2 lists the TDP theory of change component; and column 3 provides some detail on the TDP theory of change for the component listed in the second column (see OPM, 2014a). Column 4 contains the indicator name (and log frame indicators are in **bold**), column 5 lists the instrument used to collect the data for the indicator and column 6 defines the indicator. Finally, column 7 contains notes on the indicator and/or data used to construct it if relevant. Col 8 suggests some disaggregations for the indicators. The programme log frame indicators are shown in bold font.

Table 18 Quantitative baseline survey indicator definitions

1. EF reference	2. TDP theory of change component	3. TDP theory of change detail	4. Indicator	5. Instrument	6. Indicator definition	7. Notes	8. Disaggregation
PUPIL LEARNING LEVELS							
Re-1	Final impact	Improved in learning English literacy, numeracy and scientific literacy for cohorts taught by selected teachers in TDP schools 2014-2019	**Mean pupil learning levels in English literacy, numeracy and scientific literacy**	Learning assessment for sampled grade 3 pupils	Mean (scaled) score in English, maths and science & technology	Learning assessment based on NERDC level 2 curriculum.	By treatment and control By state (sample size allowing) By gender (sample size allowing)
Im-9	Intermediate impact to final impact assumption	Children having the capacity to learn from improved teaching in the language of instruction (they are school ready)					
PUPIL BACKGROUND CHARACTERISTICS							
N.A.			Age (years)	Background section in the learning assessment for sampled grade 3 pupils	Self-reported age in years		By treatment and control By state (sample size allowing)
			Gender (% female)		Female pupils as a proportion of all pupils (%)		
			Language spoken at home (% of pupils)		Pupils who speak a given language (Hausa, Fulfulde, Kanuri or other) as a proportion of all pupils (%)		
TEACHER EFFECTIVENESS							
Re-3 Effe-25	Outcome Output to outcome assumption	Improved teacher subject knowledge Teachers have the basic language, subject and pedagogical skills to absorb the new knowledge and skills available from TDP	**Teacher subject knowledge in each of English, maths and science & technology**	Teacher development needs assessment (TDNA) for TDP (treatment) and control teachers and head teachers	Number of correct answers as a proportion of the maximum score on each of the TDNA English, maths and science & technology components (%)		By treatment and control By state (sample size allowing)

1. EF reference	2. TDP theory of change component	3. TDP theory of change detail	4. Indicator	5. Instrument	6. Indicator definition	7. Notes	8. Disaggregation
Re-2	Intermediate impact	Improved teacher effectiveness in classroom	**Time teacher involves pupils in positive interaction during lesson (% of total lesson time)**	Classroom observation of sampled teachers and head teachers who teach	See Section 3.3.3 above.		By gender (sample size allowing) By state (sample size allowing)
Effe-2	Intermediate impact	Improved teacher effectiveness in classroom	**Average daily absenteeism (% of teachers absent)**	School records for all teachers at the school	Total number of teachers absent over the previous five school days divided by the total number of teachers employed over the previous five school days multiplied by 100.		
N.A.	Intermediate impact	Improved teacher effectiveness outside classroom	Teacher's knowledge to assess and monitor pupil academic progress	Teacher development needs assessment (TDNA) for TDP and control teachers and head teachers	Number of correct answers as a proportion of the maximum score on the TDNA assessing and monitoring pupil academic progress component (%)		By treatment and control By state (sample size allowing)
N.A.			Praise more than reprimand (% of teachers and head teachers who teach)		Number of teachers who use praise more frequently than reprimands as a proportion of TDP and control teachers and head teachers who teach (%)		
N.A.			End of lesson teaching practices: summarised day's lesson	Classroom observation of sampled teachers and head teachers who teach	Number of teachers who demonstrated each of the end of lesson teaching practices as a proportion of TDP and control teachers and head teachers who teach (%)	Lessons were observed for up to 36 minutes (the standard length is 35 minutes) but 30% of lessons lasted longer. For these lessons data on end of lesson practices are not available.	
N.A.			End of lesson teaching practices: revisited lesson's objectives				
N.A.			End of lesson teaching practices: gave homework				
N.A.			Teacher motivation	Teacher interview	See (Cameron, 2015) for discussion on methods and analysis		
TEACHER BACKGROUND CHARACTERISTICS							
Effe-16	Outcome to intermediate impact assumption	Selected teachers being class ready, in other words have the capacity to apply their new knowledge,	Holds NCE qualification (%) Received in-service training during last two years (%)	Teacher interview	Teachers who hold an NCE qualification as a proportion of TDP and control teachers (%) Teachers who received in-service training during the last two school years as a proportion of		By treatment and control By state (sample size allowing)

1. EF reference	2. TDP theory of change component	3. TDP theory of change detail	4. Indicator	5. Instrument	6. Indicator definition	7. Notes	8. Disaggregation
Re-20	Outcome to intermediate impact assumption	Selected teachers being retained in schools where the TDP is operating,	Experience in current school (years)		TDP and control teachers (%) Number of years teacher has worked in current school		
			Age (years)		Age in years		
N.A.			Gender (% female)		Number of female teachers as a proportion of all interviewed teachers (%)		
			Experience teaching (years)		Total number of years working as a teacher		
SCHOOL LEADERSHIP AND MANAGEMENT							
Effe-17	Outcome to intermediate impact assumption	Head teachers being motivated to lead and manage teachers well Selected teachers being supported to apply their new knowledge	Head teachers who hold formal meetings with teachers once a week or more often (%) Head teacher carried out lesson observations during the last two weeks (%) Head teacher took action to reduce pupil absenteeism last year (%) Head teacher took action to reduce teacher absenteeism last year (%)	Head teacher interview	Number of head teachers who report holding formal meetings with all or a group of teachers once a week or more often as a proportion of all head teachers (%) Number of head teachers who reported carrying out lesson observations during the last ten working days as a proportion of all head teachers (%) Number of head teachers who reported taking action to reduce pupil absenteeism during the last school year as a proportion of all head teachers (%) Number of head teachers who reported taking action to reduce teacher absenteeism during the last school year as a proportion of all head teachers (%)		By treatment and control By state (sample size allowing)
Effi-8	Output	Provide continuous support to teachers for a prolonged period of time and embed this mechanism in both schools and the TDP states' teacher education systems,	Frequency of school visits last school year: more than three times a month, two to three times a month, once a month or less	Head teacher interview	Number of head teachers reporting that a supervisor visited the school during the last school year (more than three times, two to three times or once a month or less) as a proportion of all head teachers (%)		
N.A.			SBMC exists (% of schools)		Number of schools with a school-		

1. EF reference	2. TDP theory of change component	3. TDP theory of change detail	4. Indicator	5. Instrument	6. Indicator definition	7. Notes	8. Disaggregation
					based management committee (SBMC) as a proportion of all schools (%)		
			SBMC met this term or during the preceding vacation (% of schools)		Number of schools with SBMCs where SBMCs met in the current term or preceding vacation, as a percentage of schools with SBMCs		
Effe-14	Outcome to intermediate impact assumption	Selected teachers being sufficiently extrinsically motivated to apply their new knowledge	Receipt of salary (% of teachers): always on time, usually on time, usually delayed, always delayed	Teacher interview	Number of teachers who report receiving their salary always on time, usually on time, usually delayed or always delayed as a proportion of TDP and control teachers (%)		
HEAD TEACHER BACKGROUND CHARACTERISTICS							
N.A.			Age (years)	Head teacher interview	Age in years		By treatment and control By state (sample size allowing)
			Gender (% female)		Number of female teachers as a proportion of all interviewed teachers (%)		
			Experience teaching (years)		Number of years working as a formal teacher including as a head teacher		
			Experience as head teacher (years)		Total number of years working as a head teacher		
			Experience as head teacher in current school (years)		Number of years working as head teacher at the current school		
			Holds NCE qualification (%)		Head teachers who hold an NCE qualification as a proportion of all head teachers (%)		
			Received in-service training last two years (%)		Head teachers who received in-service training during the last two school years as a proportion of all head teachers (%)		
			Receipt of salary (% of head teachers): always on time, usually on time, usually delayed, always delayed		Number of head teachers who report receiving their salary always on time, usually on time, usually delayed or always delayed as a proportion of all head teachers (%)		
SCHOOL CHARACTERISTICS							

1. EF reference	2. TDP theory of change component	3. TDP theory of change detail	4. Indicator	5. Instrument	6. Indicator definition	7. Notes	8. Disaggregation
Re-16	Intermediate impact to final impact assumption	A class size small enough to allow improved teacher effectiveness to have an impact;	Pupil-teacher ratio (PTR)	Head teacher interview and school records	Total number of grade 1-6 pupils registered at the school divided by the total number of grades 1-6 teachers employed at the school		By treatment and control By state (sample size allowing)
N.A.			Number of teachers employed (grades 1-6)		Total number of grades 1-6 teachers employed at the school		
			Number of pupils registered (grades 1-6)		Total number of grades 1-6 pupils registered at the school		
			Schools receiving support from organisations/programmes (% of all schools)		Number of schools where head teachers report that school currently receives support in cash or in-kind from organisation or programmes (such as NGOs, mosques, foreign projects, GEP, ESSPIN and private) as a proportion of all schools (%)		
			Major repair needed (% of schools)		Number of head teacher that reports that major repairs are needed as a proportion of all head teachers (%)		
			Schools with (regular or irregular) electricity supply (% of schools)		Number of schools that have an electricity supply (regular or irregular) as a proportion of all schools (%)		
N.A			Number of pupils per observed classroom (class size)	Classroom observation	Number of pupils present during classroom observation		
			Number of resources used		Number of different resources used by TDP and control teachers and head teachers who teach during the observed lessons (%)		
			Teachers who used each of: textbook, blackboard, chalk, poster, chart & pictures, improvised materials made by teacher, resources from the local environment, audio, video, science equipment, other equipment and none of the above		Number of teachers who used each type of resource during the observed lessons as a proportion of TDP and control teachers and head teachers who teach (%)		
OWNERSHIP AND USE OF SMARTPHONES							
Effe-23	Output to outcome	Teachers can access and use the audio-visual materials (i.e. the technology	Proportion of teachers and head teachers who own a mobile phone with	Teacher interview and head teacher interview	Number of teachers and head teachers who own a working		By treatment and control

1. EF reference	2. TDP theory of change component	3. TDP theory of change detail	4. Indicator	5. Instrument	6. Indicator definition	7. Notes	8. Disaggregation
	assumption	works, can be charged, is not lost, stolen or broken, is upgraded or fixed where appropriate, can be understood, etc.)	video and audio (%)		mobile phone that can use video and audio as a proportion of all teachers and head teachers (%)		By state (sample size allowing)

A.4 Qualitative evaluation matrix (as proposed by the qualitative baseline survey concept note)

Table 19 TDP in-service training qualitative baseline survey

Logframe area	Hypotheses	Assumptions	Core areas to probe in this qual. study (round 1)	Source of information
Impact: Improved learning				
Intermediate impact: improved teacher effectiveness in classroom	Improved teacher effectiveness in classroom will lead to improved learning outcomes of pupils	• Pupils **attend** lessons at school • Pupils have sufficient **nutrition** and rest to facilitate concentration during lessons • Pupils have **time** outside of school hours to reinforce learning (e.g. through discussion and homework) • **Parents encourage pupils** to learn by reinforcing positive behaviour and learning outcomes outside school • **Gender** and **economic inequalities** of pupils are identified and their impact on learning outcomes is mitigated by teachers • (Head) teachers have **feedback mechanisms** that enable them to identify improved learning outcomes and reinforce positive teaching and management practices • Differences between **language**	• What **motivates** children? • Why don't children **attend** lessons at schools? • How do pupils **behave** during lessons? • How might other obligations outside of school (**caring** for family members, household **chores**, **income generating activities**) limit pupils' school attendance or time spent on learning outside of school? • What support do **parents** provide to pupils? • Why are learning outcomes often lower amongst **girls** and children from the **poorest households**? • How do (head) teachers **monitor** pupils' learning outcomes on an ongoing basis and how does	• Pupil FGD • Head teacher KII • Teacher FGD • Informal discussion with community members/SBMC

Logframe area	Hypotheses	Assumptions	Core areas to probe in this qual. study (round 1)	Source of information
		of instruction and pupil's primary language do not prevent transmission of knowledge	this influence teacher and management practices? • How do teachers mitigate the potential negative impact of **language** differences?	
Outcome 1: Improved head teacher leadership and management	Head teachers with improved knowledge of leadership and management techniques will foster an environment in which teachers can teach effectively	• HTs are either **motivated or incentivised** to identify, incentivise and influence the positive teaching and management practices that TDP promotes • HTs **have the ability to** identify, incentivise and influence the positive teaching and management practices that TDP promotes • HTs have access to adequate support to **maintain and repair the school infrastructure and resources** that facilitate the adoption of the positive teaching practices that TDP promotes	• What **motivates** head teachers? • How do the SUBEB/LGEA and local community **hold head teachers to account**? • How do the SUBEB/LGEA and local community **facilitate HT's ability** to influence teacher behaviour? • How do head teachers communicate their view of what constitutes an 'ideal' teacher to their staff? How do head teachers **identify** the positive behaviour of such 'ideal' 'teachers? • How do head teachers **plan** for the future and how do they request the human and physical **resources** to implement this plan? • How do **teachers respond** to head teachers' attempts to influence their behaviour? • How has TDP influenced the **relationships** between teachers receiving training and other	• Head teacher KII • SUBEB KII • LGEA KII • Teacher KIIs Teacher case study

Logframe area	Hypotheses	Assumptions	Core areas to probe in this qual. study (round 1)	Source of information
			teachers?	
Outcome 2: Improved teacher subject content knowledge	Teachers with improved subject content knowledge will teach more effectively in the classroom	• Teachers **attend** school to teach** • Teachers **do not curtail** lesson duration to shorter than official timetable mandates** • Teachers are EITHER **motivated** OR **incentivised** to apply their new subject knowledge** • Teachers have access to **sufficient classroom materials** to apply new subject knowledge (*Output 3*) • Large **class sizes** do not prevent transmission of new subject knowledge** • **Curriculum design** and **textbooks** allow for new subject knowledge to be incorporated into lesson plans	• What **motivates** teachers? Why are NCE-qualified teachers more motivated than unqualified ones? • Why do teachers not attend school and how do they mitigate the negative impact of **out-of-school obligations** on their school attendance (if at all)? • Why do teachers curtail **lesson duration** and how do head teachers address this (if at all)? • What **incentive structure** do teachers face, and how does this influence their teaching practice? • How do the head teacher and **pupils' parents** hold teachers to account? • Why does a lack of **school infrastructure** (classrooms, furniture, sanitation) present a barrier to application of new subject knowledge, and how do teachers mitigate the impact of this limitation? • Why does a lack of **classroom materials** (textbooks, aids) present a barrier to new subject knowledge, and how do teachers mitigate the impact of this limitation?	• Teacher FGD • Teacher case study + photo exercise • Teacher facilitator KII • Head teacher KII • LGEA/SUBEB KII • Informal discussion with community members/SBMC

Logframe area	Hypotheses	Assumptions	Core areas to probe in this qual. study (round 1)	Source of information
Outcome 3: Improved teacher pedagogical knowledge	Teachers with improved pedagogical knowledge will teach more effectively in the classroom	• Teachers **attend** school to teach • Teachers **do not curtail** lesson duration to shorter than official timetable mandates • Teachers are EITHER **motivated** OR **incentivised** to apply their new knowledge by changing teaching practices • **Social norms** and **power relations** do not constrain the adoption of new teaching practices • Teachers have access to **sufficient classroom materials** to apply new pedagogical knowledge (*Output 3*) • Large **class sizes** do not prevent adoption of new teaching practices • **Curriculum design** and **textbooks** do not prevent adoption of new teaching practices	• What **motivates** teachers? Why are NCE-qualified teachers more motivated than unqualified ones? • Why do teachers not attend school and how do they mitigate the negative impact of **out-of-school obligations** on their school attendance (if at all)? • Why do teachers curtail **lesson duration** and how do head teachers address this (if at all)? • What **incentive structure** do teachers face, and how does this influence their teaching practice? • How do the **head teacher, school inspectors** and **pupils' parents** hold teachers to account? • To what extent do teachers feel they are well **supported** by the head teacher in school and out-of-school? • Why does a lack of **school infrastructure** (classrooms, furniture, sanitation) present a barrier to application of new teaching techniques, and how do teachers mitigate the impact of this limitation? • Why does a lack of **classroom materials** (textbooks, aids) present	• Teacher FGD • Teacher case study + photo exercise • Teacher facilitator KII • Head teacher KII • LGEA/SUBEB KII • Informal discussion with community members/SBMC

Logframe area	Hypotheses	Assumptions	Core areas to probe in this qual. study (round 1)	Source of information
			a barrier to application of new teaching techniques, and how do teachers mitigate the impact of this limitation? • What do teachers feel about the appropriateness and flexibility of the **curriculum**?	
Output 1: Collaboration and partnership	TDP cluster-based delivery model will facilitate peer-to-peer learning to improve (head) teachers' subject content and pedagogical knowledge	• Peer-to-peer model is an **effective mechanism** for transmitting information and skills when implemented as intended • Cluster meetings are facilitated in a way that **peer-to-peer learning model is implemented** • Teachers are EITHER **motivated** OR **incentivised** to attend cluster meetings and participate in peer-to-learning process • **Social norms** and **power relations** do not prevent active participation in cluster meetings by some groups (women, junior teachers, non-local teachers) • **Other obligations** (family, community, second jobs) do not prevent teachers from attending cluster meetings • **Additional forums** exist that facilitate peer-to-peer learning in an informal, less structured manner (besides cluster meetings)	• Why do teachers (not) **attend** cluster meetings? • What do teachers **discuss** during cluster meetings? • How **free and comfortable** do teachers feel to speak up during cluster meetings? • How do **gender** or other factors influence participation in cluster meetings? • To what extent do teachers share knowledge in **informal, less structured contexts**?	• Teacher FGD • Teacher facilitator KII / cluster meeting observation • Head teacher KII

Logframe area	Hypotheses	Assumptions	Core areas to probe in this qual. study (round 1)	Source of information
Output 2: Training and support	TDP training and support improves teachers' subject content and pedagogical knowledge and head teachers' knowledge of leadership and management techniques.	• Training and support is **delivered** in a way that effectively transmits information and skills to trainees • Head teachers are EITHER **motivated** OR **incentivised** to participate in learning process • Teachers are EITHER **motivated** OR **incentivise**d to participate in learning process • Teachers have **feedback mechanisms** that enables them to reinforce positive learning and correct mistakes • **Other obligations** (family, community, second jobs) do not prevent teachers from spending time using 'teacher in the pocket' resources outside of school	• What **motivates** (head) teachers? • What **incentive structure** do (head) teachers face, and how does this influence their willingness to learn? • What do (head) teachers believe constitutes **an 'effective' (head) teacher**? • How do (head) teachers receive **feedback** on their learning that enables them to know when they are learning correctly? What role do head teachers, teacher facilitators, and the learning resources themselves play in this? • Why do teachers not participate in training opportunities and how do they mitigate the negative impact of **out-of-school obligations** on the time they spend training (if at all)?	• Teacher FGD • Teacher facilitator KII • Head teacher KII
Output 3: Materials development	Teachers with access to TDP teaching materials will use them as part of more effective classroom teaching behaviours	• Teachers are EITHER **motivated** OR **incentivised** to incorporate use of new teaching materials into lessons • Teachers combine receipt of new materials with pedagogical knowledge to **use materials as encouraged by training resources**.	• To what extent do head teachers **encourage** use of TDP teaching materials? • How do teachers **feel** about the new teaching materials? Do they make **links** between newly acquired pedagogical knowledge and new teaching materials?	• Teacher FGD • Head teacher KII
Output 4: Technology	Teachers with access to	• Audio-visual resource content is	• How do teachers **feel** about	• Teacher FGD

Logframe area	Hypotheses	Assumptions	Core areas to probe in this qual. study (round 1)	Source of information
use and management	TDP 'teacher in the pocket' audio-visual resources will use these to improve their subject content and pedagogical knowledge	**relevant** to existing knowledge gaps of teachers • Audio-visual resources are **delivered** in a way that effectively transmits information and skills to trainees • Teachers are EITHER **motivated** OR **incentivised** to participate in learning process • Teachers have **feedback mechanisms** that enables them to reinforce positive learning and correct mistakes • **Other obligations** (family, community, second jobs) do not prevent teachers from spending time using 'teacher in the pocket' resources outside of school	the audio-visual resources? Do they make **links** between audio-visual resources and new teaching materials? • To what extent do teachers feel that the audio-visual resources are **delivered** in a way that improves their learning? • How do (head) teachers receive **feedback** on their learning that enables them to know when they are learning correctly? • Why do teachers not use the 'teacher in the pocket' resource more and how do they mitigate the negative impact of **out-of-school obligations** on the time they spend using it (if at all)?	• Head teacher KII
Outcome 5: INSET (In-Service Education Training) programme implementation	INSET leads to improved pedagogical knowledge	• INSET content is **relevant** to existing knowledge gaps of teachers • INSET is **delivered** in a way that effectively transmits information and skills to trainees • Teachers are EITHER **motivated** OR **incentivised** to participate in learning process •	• How do teachers **feel** about the INSET? Do they make **links** between INSET and their role as a teacher? • To what extent do teachers feel that INSET is **delivered** in a way that improves their learning? • How do (head) teachers receive **feedback** on their learning that enables them to know when they are learning correctly? • Why do teachers not use the INSET resources more and how do	• Teacher FGD • Teacher facilitator KII • Head teacher KII

Logframe area	Hypotheses	Assumptions	Core areas to probe in this qual. study (round 1)	Source of information
			they mitigate the negative impact of **out-of-school obligations** on the time they spend using it (if at all)?	
Process 1: Collaboration and partnership	Teachers are able to attend cluster meetings held by TDP	• Cluster meetings are arranged by TDP at **times** when teachers can attend • **Travel** to cluster meetings is not prohibitively expensive	• To what extent do teachers feel that cluster meetings are scheduled at convenient times and locations that minimises disruption to other obligations and cost of attendance?	• TDP KII • Teacher FGD • Teacher facilitator KII / cluster meeting observation
Process 2: Training and support	Teachers receive training and support	• Adequate engagement from TFs • Cluster meetings are arranged…	• What are the roles and responsibilities of TFs? • What challenges do they face, if any, in delivering their role as TFs?	• TDP KII • Teacher FGD • Teacher facilitator KII
Process 3: Materials development	Teachers receive teaching materials that can be used in a classroom	• TDP has **distributed teaching materials** that can be used in a classroom	• Have teachers received teaching materials? If not, why not?	• TDP KII • Teacher FGD
Process 4: Technology use and management	Teachers receive and are able to use 'teacher in the pocket' audio-visual resources	• TDP has **distributed audio-visual resources** to teachers • Teachers are adequately trained on **how to use** hardware and software • Teachers have access to **electricity** to charge hardware • **Charging** hardware is not prohibitively expensive	• Have teachers received 'teacher in the pocket' resources? If not, why not?	• TDP KII • Teacher FGD • Teacher facilitator KII /
Process 5: INSET programme implementation	Teachers are being trained through INSET	• INSET training is **arranged** by TDP • Teachers are **aware** of INSET training	• Have teachers received INSET? If not, why not?	• TDP KII • Teacher FGD • Teacher facilitator

Logframe area	Hypotheses	Assumptions	Core areas to probe in this qual. study (round 1)	Source of information
				KII
↓				
Themes in Evaluation Matrix Themes used to structure qualitative evaluation tools				

We have grouped 'assumptions' into following themes:

- Motivation
- Monitoring and Accountability (within school)
- Monitoring and Accountability (outside of school)
- Peer relationships
- Infrastructure
- Curriculum
- Relationship with pupils
- External pressures

Not all tools cover all themes.

Annex B TDP intervention factsheet for in-service training component

The purpose of the intervention factsheet is to help define the intervention boundaries and describe conditioning factors for the evaluation design by articulating aspects of the intervention design. The factsheet is meant to be a concise factual representation—without evaluative judgement—reflecting current thinking by the implementing partners.

Table 20 Factsheet for in-service training component

Intervention	In-service training	Source
Current alignment with overall programme design		
Output Logframe	TDP Output 1: Improved in-service training of primary and junior secondary school teachers	TDP Logframe, May 2014
Output indicators:	a) Number of teachers trained in the year b) Number of head teachers trained in management and development of staff (person days) c) Number of 'TFs' trained to support teachers' in-service training: this role provides the support package to teachers	TDP Logframe, May 2014
Budget allocated	DFID: £24.9 million (60%), State govts.: £16.4 million (40%)	TDP VFM Strategy Final, February 2014
Impact weighting	80%	TDP logframe, May 2014
Intervention logic		
Objectives	• Establish a school-based, cost-effective in-service training programme supported by the 'trainer in the pocket' model and similar approaches • Establish permanent cadres of teacher trainers responsible for the coordination, development and delivery of in-service programmes • Assist states to provide regular and ongoing training and support to a total of 62,000 teachers by 2019	TDP INSET Strategy, January 2014
Intended outcomes	• More effective teachers in target schools • More effective teacher educators in colleges	TDP Logframe, May 2014
Outcome indicators	• Percentage change (year on year) in the proportion of time teachers involve students in positive interactions in a lesson • Change in teacher absenteeism as a percentage of the average days absent to the number or contracted working days measured year on year • Percentage change (year on year) in teacher subject and pedagogic knowledge competency	TDP Logframe, May 2014
Intended impact	• Improved student learning in target schools	TDP Logframe, May 2014
Impact indicators	• Percentage change (year on year) in student learning outcomes: +3 percentage points in each of English, maths and science at the end of Phase 1 (measured through endline in Year 3) and another +3 percentage points at the end line of Phase 2 (Year 5). Baseline to be established in October 2014 (through a school survey)	TDP Logframe, May 2014
Beneficiaries	• Ultimate beneficiaries: students in Primary (P) P1–6 and JSS 1–3 • Primary beneficiaries: teachers in P1–6 and JSS 1–3 of English (31,000), maths (31,000), science and technology (31,000) • Secondary beneficiaries: state-based teacher trainers (Teacher	TDP INSET Strategy, Jan 2014

Intervention	In-service training	Source
	Development Teams) and school supervisors (TFs)	
Scope of intervention		
Target population	**Whole programme:** The in-service training of teachers in the classroom will reach 62,000 teachers, who will be trained over three years each in English, maths and science and technology. In turn, for every year they continue as teachers, they will improve the learning outcomes of over 2 million students in primary and junior secondary schools. **Pilot:** TDP will reach 2,000 teachers from 500 schools across selected LGEAs in three states by 2015 as part of its pilot plan. TDP's intervention in the next phase states (Kano, Kaduna and Niger) will start from 2016 and will continue till the end of the programme.	
Geographical scope	**Whole programme:** Northern Nigeria, in six states over the course of the six years. **Pilot:** In the first two years, otherwise referred to as the first phase/pilot, the programme will be implemented in the states of Zamfara, Katsina, and Jigawa. In the second phase, the programme will be implemented in Kano, Kaduna and Niger State.	
Implementation timeline	TDP will reach 2,000 teachers from 500 schools across selected LGEAs in three states by 2015, as part of its pilot plan. TDP intervention in the next phase states (Kano, Kaduna and Niger) will start from 2016 and will continue till the end of the programme.	
Intervention components		
School-based interventions	'New classroom activities' will be at the core of all programme activities, i.e. – teacher training, materials for students and teachers, and teacher support.	TDP INSET Strategy, January 2014
'Trainer in the pocket':	Access to audio-visual resources, anytime and anywhere, ensured through the use of mobile technology.	TDP INSET Strategy, January 2014
Continuous support	Ensured continuous support to teachers over several years through multiple layers and mechanisms (workshops, cluster meetings, classroom observation, peer support, self-study materials etc.) instead of one-off training. The support mechanism for continuous professional development will be institutionalised within the school (Teacher Development Team) and state (TFs).	TDP INSET Strategy, January 2014
Intervention details		
Partnership with other DFID Nigeria education programmes	Joint activities with ESSPIN, GEP. Partner with SUBEB as the institutional home for TDP INSET activities, with the Director Training being the focal point for the programme.	
Build and embed a whole school development approach	By gradually training and supporting a maximum of six teachers per school in English, maths and science and technology. Provide continuous support to teachers for a prolonged period of time through different mechanisms, and embed similar support mechanism into the school culture as well as state's teacher education system.	
Collaboration with ESSPIN and GEP with regard to head teacher activities	Engage the head teachers actively in the pedagogy, as well as leadership and management training.	
Selection of LGAs, schools and teachers	Support states to identify the LGAs, schools and teachers to participate in the pilot phase of TDP.	
Recruitment	Teacher Development Team (at state level) and the TFs (at LGEA and cluster level). In consultation with SUBEB and CoEs, recruit a pool of teacher trainers as the Teacher Development Team based in the state. Recruit TFs from the current school supervisor cadre (School Support Officers/Quality Assurance	

Intervention	In-service training	Source
	Officers (QAOs)) at LGEA level, based on a rigorous selection process.	
Material development	• In the area of materials development, the plan is to, in the first instance, benefit from and adapt materials from existing programmes – lesson plans in English and maths from ESSPIN and head teacher training materials from both ESSPIN and GEP 3. • TDP will then also develop audio-visual materials for use by all the beneficiary groups of the Teacher Development Team, the TFs and teachers. The plan is to develop and pilot the materials in phases, beginning with P1–3 English and maths, then P4–6 English and maths, then P4–6 science and technology, then JSS 1–3 English, maths and science and technology. • The programme is proposing an approach whereby English is the language of instruction in the audio-visual and print materials for P4–6 upwards but Hausa is used for maths (only) materials in P1–3.	
Innovation and technology	The programme aims to undertake a technology assessment study to decide on the final options for the mobile technology to support teachers. State governments have communicated their interest in supporting the procurement of technology tools to aid teacher training, although firm commitments backed up with budgets have not yet been put in place.	
Quality assurance	The in-service strategy is predicated on quality assurance of the component which incorporates oversight by the school and the education administration at local and state levels.	
Intervention stakeholders		
Implementing partners	DFID (through Mott MacDonald) State in three focal states: SUBEBs Federal: NCCE	TDP Annual Review Final, March 2014; TDP INSET Strategy, January 2014
Evidence partners	EDOREN	
Funding partners	DFID, state governments	
Intervention-specific monitoring and evaluation (M&E) plans		
Internal	M&E strategy document.	
External	External team to implement baseline survey for in-service teacher training component and subsequent evaluations.	
Identified assumptions and risks		
State governments' contributions to TDP are made in full and on time.		TDP Logframe, May 2014
Issues around teacher workforce management start to be addressed by SUBEBs and constraints begin to be tackled effectively.		TDP Logframe, May 2014
Corruption in the education system begins to be addressed, especially in relation to teacher management issues.		TDP Logframe, May 2014
TOC's evidence based on English in Action Bangladesh How relevant is this for Nigeria? Can English in Action's success be transplanted to the Nigerian context, and to what extent?		Annual Review 2014
Directly influencing interventions		
ESSPIN	In Jigawa, ESSPIN is also running teacher/head teacher training	Revised GEP3 Logframe, 2013
GEP3	In Katsina and Zamfara, GEP3 is supporting girls' education	
USAID's P 1/2	Sokoto, Bauchi	Annual Review

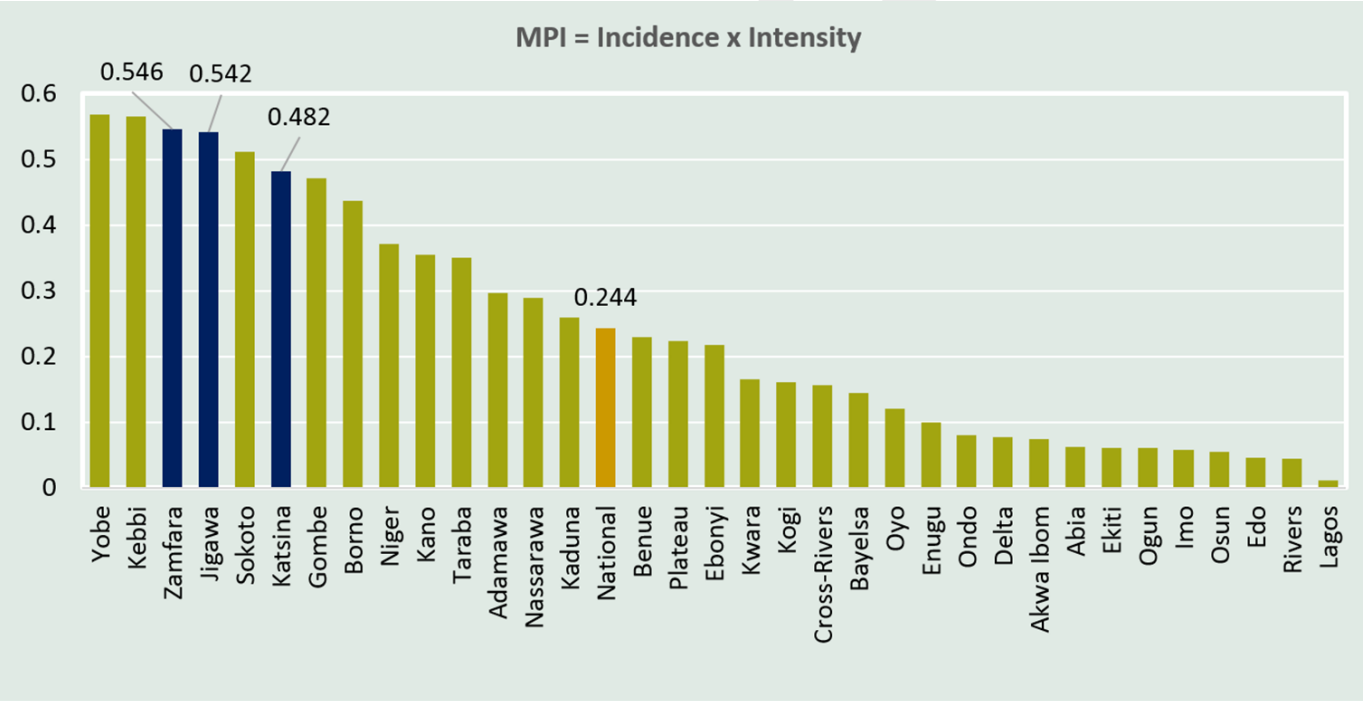
Intervention	In-service training	Source
reading skills in Hausa project		2014
Existing education administration	System of salaries, training, career progression, etc.: what determines teacher motivation and will these weaken the impact of TDP?	TDP Annual Review Final, March 2014; TDP INSET Strategy, January 2014
Political context	• CoE staff were on strike from January 2014 to the end of February 2014. Was this a one-off or will this be repeated? • In the event of federal leadership change in 2015, will counterpart contributions continue to be met? • Will there be 'slippage' in meeting programme milestones as a result of the general elections?	TDP Annual Review Final, March 2014; TDP INSET Strategy, January 2014
Reference documents		
Intervention documents	TDP Logframe, May 2014 TDP INSET Strategy, January 2014	
Evidence documents (studies, evaluations)	TDP Annual Review Final, March 2014 TDP Evaluation of First and Ongoing Activities, January 2014	

Annex C Contextual profiles and state maps of TDP Phase 1 states

C.1 TDP's operational context includes some of the 'poorest' states in Nigeria

TDP's Phase 1 states (Zamfara, Jigawa and Katsina) are among the poorest states in Nigeria. In 2010, 70%–75% of the population in these states had a purchasing power parity adjusted income of less than $1 a day, making them the states with the highest incidence of income poverty in Nigeria (Nigerian Bureau of Statistics 2012, 9). These states also fare poorly on non-income measures of poverty. The Multidimensional Poverty Index, which measures poverty based on the incidence and intensity of deprivation in education, health and living standards, found that Zamfara, Jigawa and Katsina were among the third, fourth and sixth poorest states in Nigeria (Figure 5) (Oxford Poverty and Human Development Initiative, 2014).

Figure 5 Multidimensional poverty indices for Nigerian states



Source: OPHI, 2014. Note: Data excludes Anambra and Bauchi States

C.2 Conflict in northern Nigeria

Despite intense international attention on the Boko Haram insurgency in northern Nigeria, and the declaration of a state of emergency in neighbouring Yobe state since May 2013, Jigawa is often considered one of Nigeria's least violent states. It was the country's second-least violent state between 2009 and 2013, on a per capita basis (Taft and Haken 2015, 60). There has been some spillover of the conflict with Boko Haram: for example, in May 2014 when suspected Boko Haram members shut down over 30 primary and secondary schools in Gwaram LGA (Sahara Reporters 2014). Katsina has also largely escaped the Boko Haram insurgency in northern Nigeria. It was the country's third-least violent state between 2009 and 2013 (Taft and Haken 2015, 61), although tensions between pastoralists and settled communities have sometimes spilled over into violence (Nigeria Watch, 2015. Zamfara similarly has largely avoided entanglement in the Boko Haram insurgency in northern Nigeria. It was the country's fifth-least violent state between 2009 and

2013 on a per capita basis; however, levels of violence have increased significantly since 2011 (Taft and Haken 2015, 94). This has been driven in part by rising tensions between Fulani pastoralists and settled communities.

Box 4 summarises key findings on education, conflict and violence in northern Nigeria based on research by ESSPIN (2014).

Box 4 Conflict and violence in northern Nigeria

- Violence in the neighbouring state of Yobe has caused an influx of displaced people to Jigawa, seeking safety, which may potentially strain services provision, including education.
- Jigawa's geographic terrain requires pupils to travel long distances through semi-forested pathways to get to school, and this heightens the risk of pupils' abduction and sexual violence. This is likely to affect pupils' attendance at school and is reported to have increased the rate of drop out, especially among girls.
- There have been some incidents of conflict and violence in Jigawa that have caused pupils and teachers to stay away from school or that have even resulted in periods of school closure. For example, a bomb was planted close to a primary school in February 2014. Even where schools are not directly attacked, news over the radio of attacks in other communities can cause fear and can keep pupils away from school.
- One of the common themes encountered in ESSPIN's research was a perceived incompatibility between 'Western education' and Islam. The belief that 'western education destroys the fabric of their society' was recorded in 60% of the participatory rural appraisals carried out as part of the ESSPIN research in Kaduna, Kano and Jigawa.
- All groups interviewed as part of the participatory rural appraisals identified the 2011 election as a recent example of conflict or violence in their communities. In contrast to neighbouring states, Jigawa mostly avoided escalation into widespread violence in 2011.

Source: ESSPIN (2014).

Figure 6, Figure 7 and Figure 8 below show the geographical coverage of TDP LGAs in the three Phase 1 states.

Figure 6 LGAs in Jigawa with TDP treatment and control schools

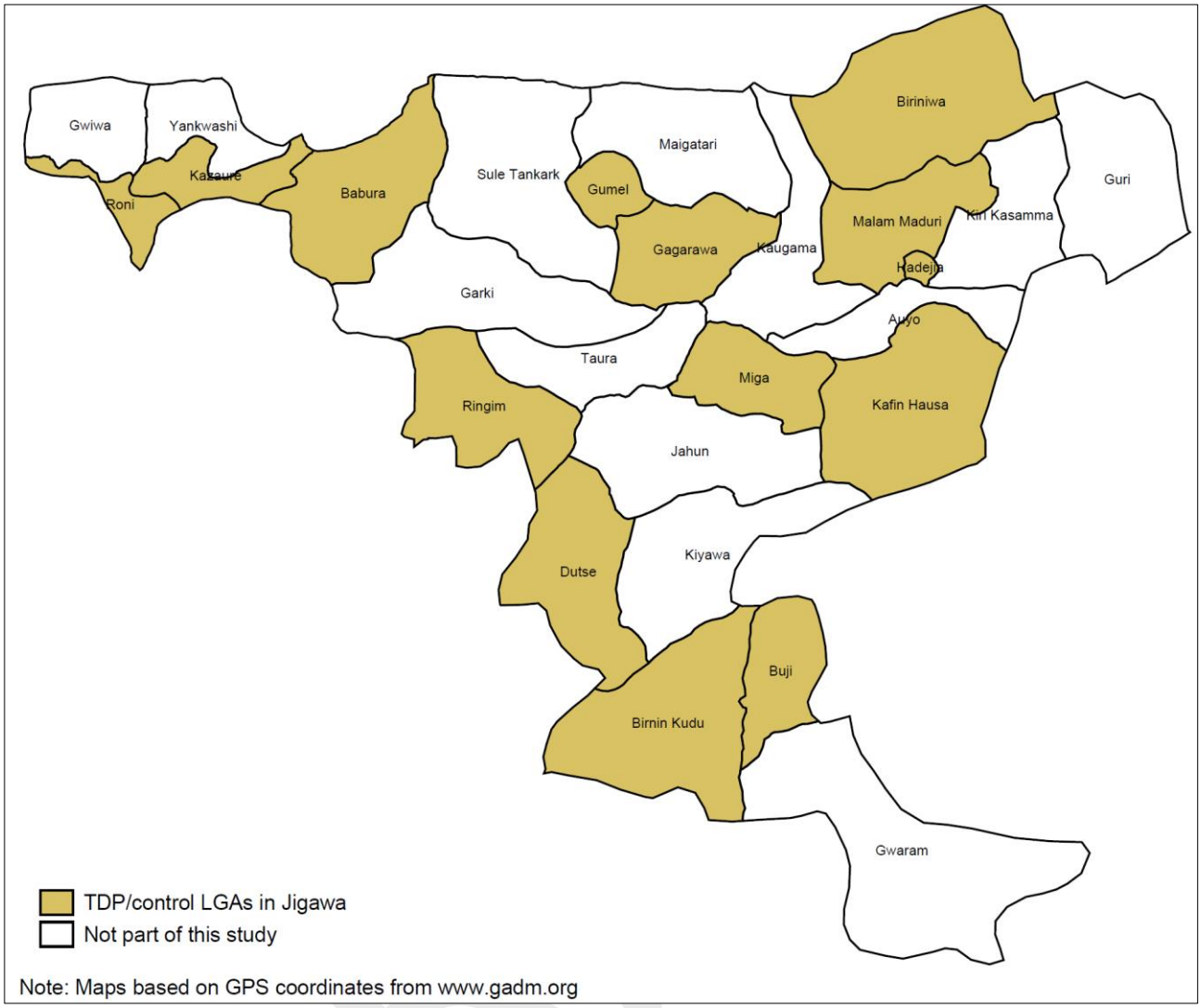


Figure 7 LGAs in Katsina with TDP treatment and control schools

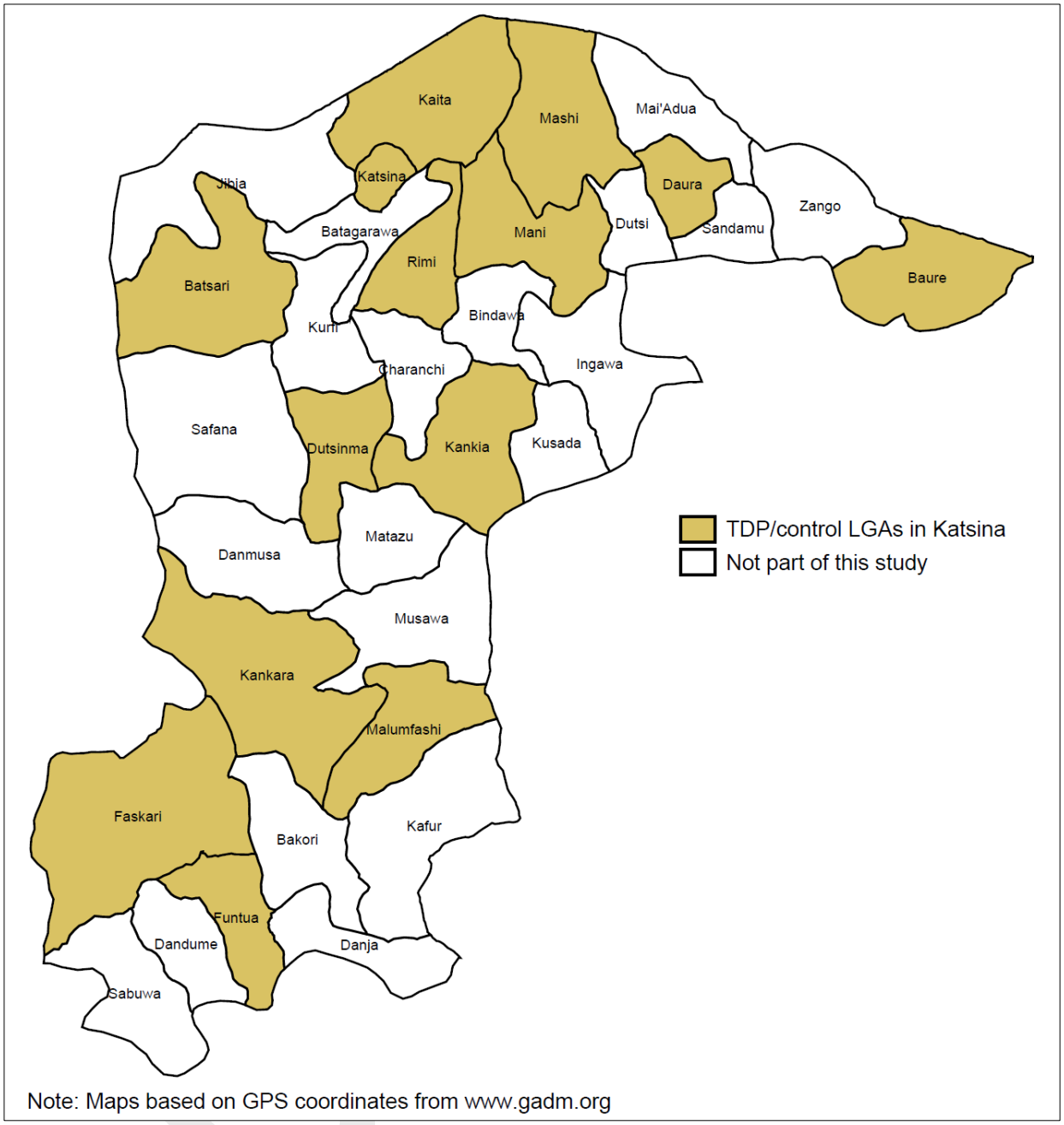
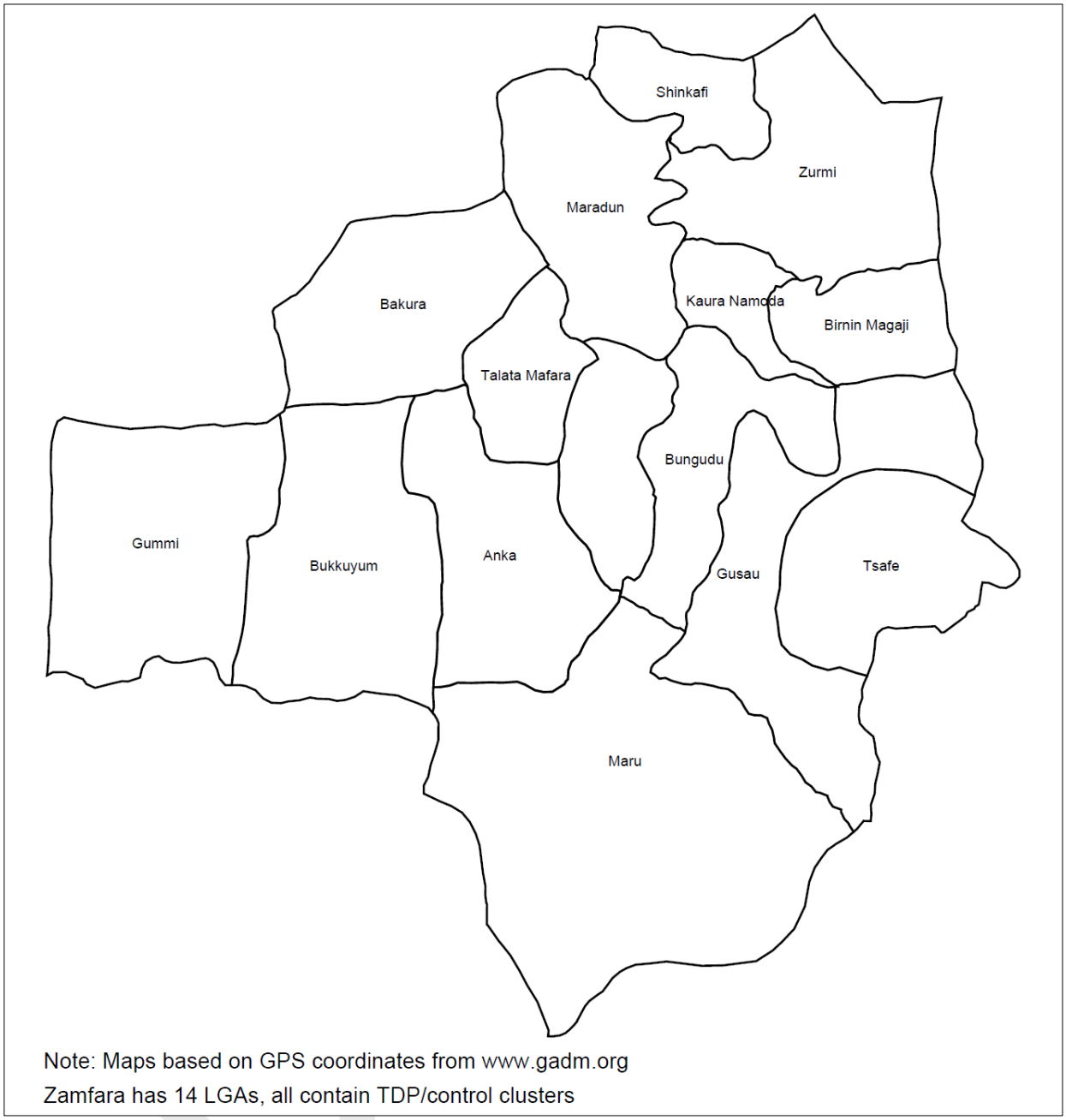


Figure 8 LGAs in Zamfara with TDP treatment and control schools[37]



[37] The blank area in the middle of the map is part of Maru local government.

Annex D Final sample design and weighting procedures[38]

The TDP school baseline survey design takes into account the implementation plans for TDP and the indicators in the project logframe. The final plans for the TDP programme have evolved over time, and the final sample design for the baseline survey has been modified accordingly.

This note begins by briefly describing the implementation of the TDP programme and the final sampling plan for the baseline survey, followed by the weighting procedures based on that sample design. A useful reference document is the earlier sampling note written for the TDP evaluation framework (EDOREN, 2014) titled 'Description of Methodology for Calculating the Minimum Detectable Effect (MDE) for a Difference of Differences Estimate'.

D.1 Panelling pupils and teachers

There will be a TDP endline survey in 2018, after three years of the programme implementation. For this reason the baseline sample schools will be part of a panel that will be followed up in the endline survey to measure trends in the indicators. The sample pupils (eight per school) and teachers (one head teacher and three other teachers chosen for the programme) will also be panelled within each school in the treatment and control cluster. A sample of control schools was also included in the baseline survey, so that the trends in the key indicators for TDP schools can be compared to those for the control group. This will involve a 'difference-of-differences' analysis, as described in the reference documents.

D.2 Implementation of TDP during Phase 1 and the population for this evaluation study

During the first phase the TDP was implemented in public schools in 14 LGAs within each of the three states (Jigawa, Katsina and Zamfara). Within each LGA the schools were clustered based on geographical proximity, in order to facilitate the training and periodic meetings of the teachers in each cluster, and to create a broader peer network within the locality. Within each LGA, two clusters of 12 primary schools each were identified: one cluster was randomly assigned to the TDP treatment group, and the other to the control group. This strategy was related to the evaluation plan for measuring the impact of the TDP intervention in the treatment schools as compared to a similar control group without the intervention. In this way a total of 42 clusters were assigned to the treatment group in the three states, with a corresponding total of 42 control clusters in the same LGAs. With 12 primary schools in each cluster, the TDP covered a total of 504 schools in the three states, and the control group also included 504 schools in these states. Within each school selected for the TDP, the Phase 1 intervention involved the training of four primary teachers: two in English and two in maths.

D.3 Experimental evaluation design

This TDP implementation was a type of constrained-RCT (random control-treatment), so the population being studied in the TDP baseline survey consists of the set 42 treatment clusters and

[38] This note has been drafted by David J. Megill (independent sampling consultant for EDOREN). The sampling and weighting procedures described in this report were developed in collaboration with various staff from EDOREN, including Sourovi De, Alex Hurrell and Matthew Powell, as well as Bukola Oyinloye of TDP. The sampling consultant appreciates their collaboration. This technical assistance was provided through the EDOREN Project, funded by DFID.

42 control clusters in the three states. Originally, each cluster had 12 primary schools, but later it was found that a few of the schools did not have eligible Grade 3 pupils, who were the subject of the pupil tests as part of the evaluation. Therefore the final population of schools for some clusters had less than 12 eligible schools. The sample for the baseline survey was selected to represent the eligible schools in the clusters for the three states. Inferences can only be made for all eligible schools in the clusters for each state. It should also be pointed out that the clusters of schools were only established in a purposive sample of 14 LGAs in each of the three states. Therefore the sample for the baseline survey was not designed to be representative at the state level.

Within the eligible schools of the treatment and control clusters in each state, the only teachers eligible to be included in the baseline survey for the treatment clusters were the three teachers receiving the TDP training and the head teacher. A similar group of four teachers was chosen in schools of the control clusters. However, some treatment schools had less than four eligible Grade 3 teachers, in which case all of them received TDP training in the treatment schools. In the case of the pupil tests, the population for the evaluation study consisted of all Grade 3 pupils who were being taught English, maths or science in the current school term by one of the teachers chosen for the study in the treatment and the control schools, and from this population, a random sample of eight pupils was drawn.

D.4 Replacement of schools

In some cases it was necessary to replace sample schools that were ineligible because they did not have any eligible third grade pupils, based on the criteria described above; a few other schools were replaced because they were inaccessible or unsafe to visit. In both cases the replacement school was randomly selected from the remaining eligible schools in the same cluster. However, as specified later in the weighting procedures, it was necessary to identify the ineligible schools that were replaced. In this case any ineligible sample schools were also excluded from the total number of eligible schools in the cluster used for calculating the weights. The weighting procedures are designed to adjust the weights for any non-response of eligible schools as well as any correction to the total number of eligible schools in the cluster.

D.5 Sample stages

Once the 14 treatment clusters and 14 control clusters were established in each state, the sampling frame consisted of all the eligible public primary schools in each cluster; most clusters had 12 eligible schools each, but a few clusters had fewer schools. As mentioned above, the eligible public schools in each cluster were those with Grade 3 pupils who were taught English, maths or science by at least one of the TDP/control teachers. In this case the eligible schools in each cluster were considered the primary sampling units selected at the first sampling stage for the baseline survey.

The stratification of the sampling frame for the TDP baseline survey is by individual treatment or control cluster, since an independent sample of schools was selected from each cluster in the frame. In this case these are not 'clusters', in classic sampling terminology – actually, each primary sampling unit (school) is a cluster of teachers and pupils. Each cluster of schools is considered to be a separate stratum, and the study population consists of all the TDP and control clusters in the three states. Within each cluster stratum, the individual schools are sampling clusters of teachers and pupils.

The first sampling stage consisted of randomly selecting a sample of four schools from each of the 14 treatment clusters and 14 control clusters in each state. All of the four (or fewer) teachers who received TDP training in each sample treatment school, and the corresponding group of up to four teachers in each control school, were selected (with certainty that they would be tested and observed for the baseline survey); the head teacher from each of these sample schools was also selected.

For the pupil tests a sample of eight Grade 3 pupils was randomly selected for the TDP baseline survey from a list of all the eligible Grade 3 pupils. In the case of small schools with eight or fewer eligible Grade 3 pupils, all were selected for the baseline survey. For each sample school, the number of eligible Grade 3 pupils attending school on the day of the team visit was entered into a CAPI database, and a CAPI programme randomly selected the eight sample pupils to be tested.

D.6 Weighting procedures for TDP baseline survey

In order to make inferences from the TDP baseline survey data, appropriate weights were assigned to each sample school, teacher and pupil. The weights were equal to the inverse of the overall sampling probabilities, taking into account each stage of selection. The school, teacher and pupil weights were calculated at the school level. Based on the sample design described above, the probability and corresponding weight for the sample schools was calculated as follows:

$$p_{Sh} = \frac{n_h}{N_h} \quad \text{and} \quad W_{Sh} = \frac{N_h}{n_h},$$

Where

p_{Sh} = probability of selection for the sample schools in cluster (stratum) h

n_h = number of sample primary schools successfully enumerated in cluster h for the TDP baseline survey; generally $n_h = 4$

N_h = total number of eligible primary schools with Grade 3 pupils in cluster h; generally $N_h = 12$

W_{Sh} = weight of sample schools in cluster h

The subscript S is used to indicate that these are school-level probabilities and weights; the schools' weights are the same within each stratum (cluster). In the case of clusters in which fewer than four sample schools were successfully enumerated for the TDP baseline survey, this formula automatically adjusts the weight for non-response. In the case of a cluster where fewer than 12 schools were found to be eligible, this weight also adjusts for the smaller number of eligible schools in the cluster.

Each sample school has one head teacher, so the head teacher has the same weight as the school. Since all of the four teachers receiving TDP training in each sample treatment school and the corresponding group of four teachers chosen in each sample control school are included in the TDP baseline survey, the teacher weights are generally equal to the school weights. In the case of

small schools with fewer than four eligible teachers, the teacher weight would also be equal to the school weight if all these teachers were successfully tested and observed. However, there were a few cases of sample schools where some eligible teachers could not be enumerated, in which case it was necessary to adjust the weight for non-response. In this case the teacher weight was calculated as follows:

$$W_{Thi} = W_{Sh} \times \frac{T_{hi}}{T'_{hi}}$$

Where

W_{Thi} = weight for teachers in the i-th sample school in cluster (stratum) h

T_{hi} = number of eligible teachers included in the study for the i-th sample school in cluster h; generally $T_{hi} = 4$

T'_{hi} = number of eligible teachers with completed interviews in the i-th sample school in cluster h

In the case of classroom observations, the weights were calculated in a similar way as the teacher interview weights, but in this case T'_{hi} was the number of eligible teachers who had been successfully observed (and similarly for the TDNAs).

The weights for the sample Grade 3 pupils who were tested involves components from two sampling stages. The first component of the weight was the school weight defined previously. The second component was the inverse of the within-school probability of selection for the sample pupils. In this case the pupil weights can be defined as follows:

$$W_{Phi} = W_{Sh} \times \frac{P_{hi}}{p_{hi}},$$

Where

W_{Phi} = weight for Grade 3 sample pupils who were tested in the i-th sample school in cluster h

P_{hi} = number of eligible Grade 3 pupils in the i-th sample school in cluster h

p_{hi} = number of sample Grade 3 pupils with completed tests in the i-th sample school in cluster h; generally this is 8

Annex E School selection guidelines to SUBEBS (Jigawa version)

Thank you for participating in this school selection process. Below are some guidelines to help you choose the TDP clusters of schools.

Step 1: Select four similar clusters of schools in each LGA

The TDP will work with the already existing ESSPIN clusters; therefore please select four ESSPIN schools clusters in each LGA. Each cluster will have five to seven schools and, together, all four clusters will have 24 schools. Moreover, keep in mind the TDP requirement that chosen schools must have at least four teachers: that is, at least three teachers plus one head teacher. These teachers should be teaching at least one of English, maths or science between Grades 1 and 3.

Please ensure that the four ESSPIN school clusters in an LGA are ***as similar to each other as possible***. Similarity between schools can be based on the following (non-exhaustive) criteria:

- location of school: urban or rural;
- size of school, i.e. in terms of number of classrooms;
- size of school, i.e. in terms of number of pupils;
- approximate number of teachers;
- presence or absence of an SBMC; and
- state and/or level of infrastructure in school.

The ultimate objective is to select a set of four ESSPIN school clusters in an LGA, each with five to seven schools in it, which are similar to one another in terms of the above (or any other) criteria.

For example: if in an LGA, you choose one cluster consisting of rural schools, most of which are small in size, then it would be appropriate to find three other ESSPIN clusters of five to seven schools which also have a high number of rural schools that are small in size. To choose clusters of schools which have mostly large urban schools would not lead to similar sets of ESSPIN school clusters!

Step 2: Select teachers for TDP training in the selected schools

Once four clusters of ESSPIN schools have been selected per LGA, the LGEA Education Secretary and the head teachers of these schools are requested to select and record the names of three teachers to participate in training under TDP. Two teachers must be selected for English, and another two must be selected for mathematics. The head teacher is included among these four. The teachers selected should also have experience teaching science and technology as two out of the four teachers will be again selected to participate in the science and technology training in 2015/2016.

Step 3: Assign two clusters per LGA to the programme

Again, once the four clusters have been selected, two of them will be chosen for immediate participation in the TDP and the remaining two chosen for later participation. The two clusters for either immediate or later participation will each have a total of 12 schools. Thus, two ESSPIN clusters, totalling 12 schools, make up one TDP cluster. The choice of which two clusters to choose for immediate participation in the TDP will be made in a random process, such as a coin toss. This process will be jointly managed by the TDP and its state partners.

Thank you!

Annex F Notes on scaling the pupil test scores using Rasch analysis

F.1 Preliminaries

F.1.1 Dealing with integer scores

Since the Rasch model can only deal with integer scores, the raw scored responses were re-coded. For example, an item with scores 0, 0.5 and 1 was re-coded into scores 0, 1 and 2. Similarly, 0, 0.25, 0.50, 0.75 and 1 were re-coded into 0, 1, 2, 3 and 4. After looking at the first results, some of those categories were collapsed, either because they were almost empty or because the ability of pupils in one category was close to the ability of pupils in another category.

An example follows (Box 5). The item analysis from IRT scaling produced tables for each item. Below is the item analysis table for item q4_score, where pupils are asked to point to their nose, knee, ear, toes (based on verbal cues). The first column (*Label*) presents the re-coded scores of the original data (0=1, 0.25=1, 0.50=2, 0.75=3, 1=4). The second column shows the collapsed scores. As can be seen in the second last column, the average ability of the pupils in category four (-0.58) is very close to the ability of pupils in category three (-0.53), and also in categories two and one (-1.80 and -2.06, respectively). It is important that the average ability increases with each increasing score. Therefore, the categories are scored as [0,1,1,2,2]. The other reason for collapsing categories three and four is the small number of pupils in these categories. The third column (*Count*) gives the number of pupils in each category and the fourth column (*% of tot*) the corresponding percentages. The fifth (*Pt Bis*) and sixth (*t (p)*) give the point bi-serial for each category and the corresponding t-value and probability. The point bi-serial is the correlation between being in a category and the raw total scores on all other items. A point-biserial coefficient is a special type of correlation coefficient that relates observed item responses to a total test score. Usually the point-biserial is negative for the zero score.

Box 5 Dealing with integer values for Rasch analysis

```
item:4 (q4_score)
Cases for this item  2571  Item-Rest Cor. 0.44  Item-Total Cor. 0.54
Item Threshold(s):  -2.34 1.02  Weighted MNSQ  1.21
Item Delta(s):    -2.31 0.98
-------------------------------------------------------------------------------
 Label  Score   Count  % of tot Pt Bis   t (p)  PV1Avg:1 PV1 SD:1
-------------------------------------------------------------------------------
  0    0.00   1523   59.24  -0.39  -21.45(.000) -3.27   1.35
  1    1.00   515   20.03  0.13   6.58(.000) -2.06   1.36
  2    1.00   395   15.36  0.20  10.23(.000) -1.80   1.37
  3    2.00    62    2.41  0.21  10.62(.000) -0.53   1.47
  4    2.00    76    2.96  0.22  11.42(.000) -0.58   1.65
===============================================================================
```

F.1.2 Dealing with test items that were skipped and missing values

Since there were many skips in the test, leading to missing values, these were re-coded to 'S' for skip and scored as 0 in ConQuest, because the reason for skipping the item was that the pupils was very unlikely to be successful in the item given earlier responses. The item analysis table below (Box 6) shows item q7_score where pupils were asked to match the first letters of names of everyday objects/animals. The code 'S' is included in the first column. The second column shows

that it is scored as 0. The third column shows that 1,981 (77%) of the students skipped the item. As expected, their ability estimate in the second last column is very low (-3.10).

Box 6 Dealing with skips and missing values for Rasch analysis

```
item:7 (q7_score)
Cases for this item  2571  Item-Rest Cor. 0.65  Item-Total Cor. 0.72
Item Threshold(s):  -0.68 0.10  Weighted MNSQ  0.80
Item Delta(s):    -0.07 -0.51
-------------------------------------------------------------------------------
 Label  Score   Count  % of tot Pt Bis   t (p)  PV1Avg:1 PV1 SD:1
-------------------------------------------------------------------------------
  0    0.00   202    7.86  0.08   3.98(.000) -2.31   1.43
  1    1.00   212    8.25  0.24  12.69(.000) -1.33   1.04
  2    2.00   176    6.85  0.58  36.36(.000) 0.32   1.22
  S    0.00   1981   77.05  -0.56  -34.17(.000) -3.10   1.27
===============================================================================
```

The skipping procedure causes an issue for the Rasch model. One assumption of the Rasch model is that items are independent of each other. In other words, the probability of responding correctly to one item should not affect the probability of responding correctly to another item. Skipping items based on the response of a previous item causes dependency between items (also called *local dependence*). This violates the assumption of the Rasch model. An easy and theoretically correct way to deal with this is to merge items that show local dependence with each other into one item. In the simplest case, all pupils that respond incorrectly on Q6 will receive a score of 0 on the merged item, pupils that are successful on Q6, but not on Q7 receive a score of 1 and pupils that are successful on both items receive a score of 2. For the moment, the analysis has treated the items with empty values (the skipped items) as independent items. If time permitted, it would be better to merge an item that could be skipped with the item that preceded it.

F.1.3 Meaning of mean square fit statistics and discriminatory power of items

The expected value of the weighted mean square fit statistics (MNSQs) (the slope of the item) is equal to 1. Items with higher weighted MNSQs have flatter slopes (discriminate less well between high and low ability students) and either measure a different construct or are confusing because of some error in the item or the scoring. Items with a weighted MNSQ lower than 1 have steeper slopes. This can be a reflection of local dependence. The fact that items could be skipped depending on responses to preceding items suggests some occurrence of local dependence.

F.2 English literacy test

While some weighted MNSQs seem high (more than 1.2), the ordering of average ability across the scores looks good. These high weighted MNSQs are probably caused by the very low weighted MNSQ in the items that were skipped by many students, because the average weighted MNSQ is always approximately 1. No items need to be removed from the scale.

The item-person maps showed that the English literacy test was quite difficult for this group of pupils. The vertical plot shows the distribution of the pupils and the distribution of the items. High-performing pupils and difficult items are at the top of the scale (in logits) and low-performing pupils and easy items at the bottom. There are not many easy items for the target population. The

bottom half of the pupil distribution does not master more than two items. The Person Separation Reliability for the literacy test is 0.77 and Cronbach's alpha is 0.83. However, local dependence between items causes overestimation of the reliability and of the variance in performance.

F.3 Numeracy

Collapsing of item scores was performed for some of the items. While some weighted MNSQs seem high (more than 1.2), the ordering of average ability across the scores looks good. These high weighted MNSQs are probably caused by the very low weighted MNSQs in the items that were skipped by many students, because the average weighted MNSQ is always approximately 1. No items need to be removed from the scale. The reliability is overestimated because of the local dependence. The Person Separation Reliability for the numeracy test is 0.79 and Cronbach's alpha is 0.86. Test targeting is similar to the English literacy test: the test is, on average, difficult for the student population.

F.4 Scientific literacy

Some, but not many, categories have been collapsed. None of the items need to be removed from the scale. Skipping causes less of an issue for the science test. Targeting of the science test is good. There are some items at all ability levels. The Person Separation Reliability is 0.83 and Cronbach's alpha is 0.78.

F.5 Transformation of pupil scores

The student ability scores were estimated in logits. They were transformed to a scale with a mean of 500 and a standard deviation of 100. These values are arbitrary, but practical. If the standard deviation is 100, summarising statistics can be reported in integers (no decimals are necessary). In addition, if a standard deviation is 100, it is easy to derive effect sizes. For example, a difference of 20 score points is equal to an effect size of 0.2 (20 / 100). 500 is chosen as the mean score so it is unlikely there will be any negative achievement scores. The transformations were as follows:
$\theta_n^* = [(\theta_n - mn)/sd] * 100 + 500$ where θ_n is the ability for pupil *n* in logits and θ_n^* the transformed ability, *mn* is the weighted mean score of all pupils and *sd* is the weighted standard deviation of all pupils. The values of the means and standard deviation for each assessment domain are included in the following table. The transformations parameters for the sub-scales of English literacy and numeracy were identical to the transformation for the overall scale.

Table 21 Mean and standard deviation for each assessment domain in the pupil test

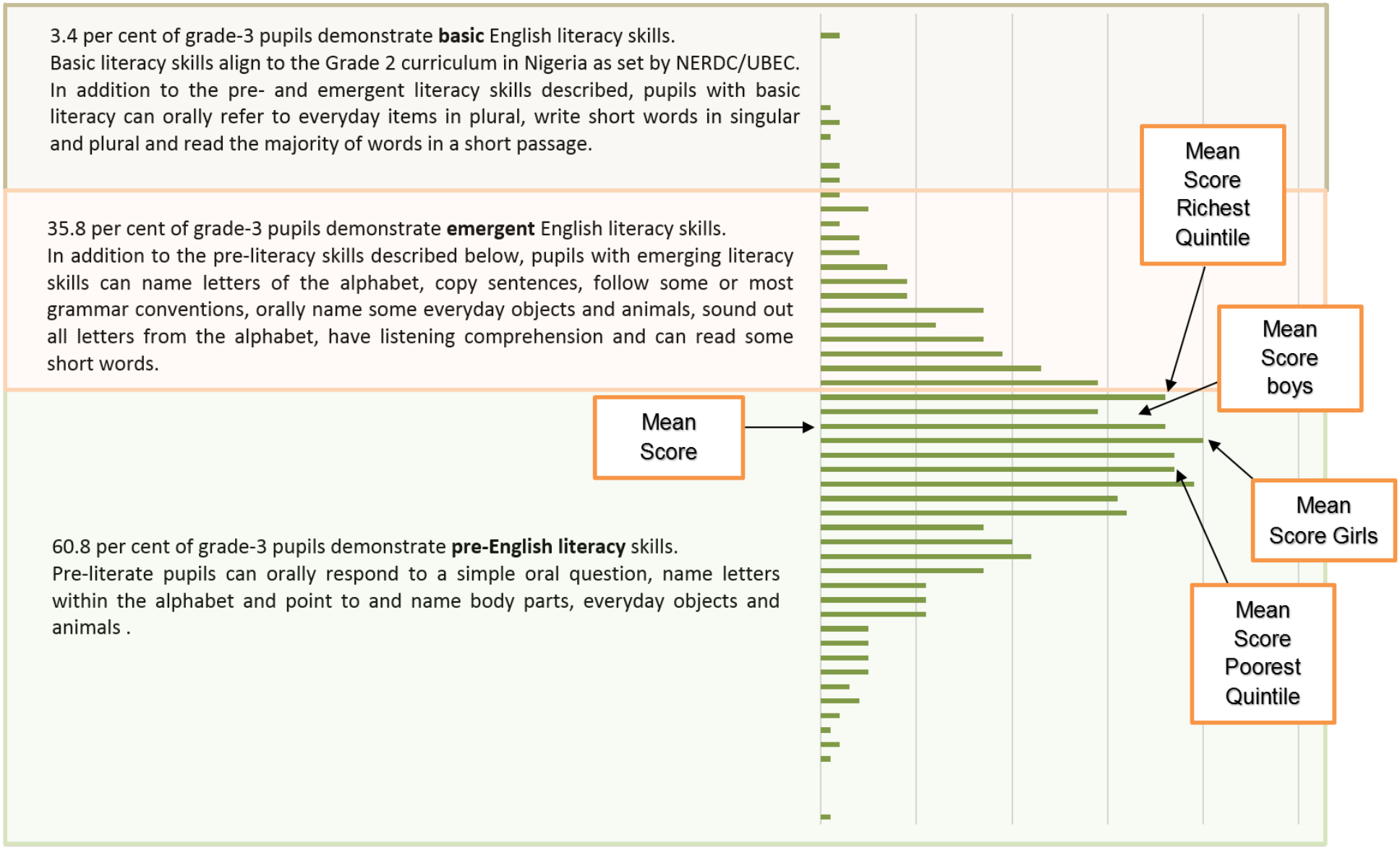
	Literacy	Numeracy	Science
Mean	-2.23	-1.63	0.82
Std. dev.	1.71	1.61	0.89

Annex G Item-person maps for English literacy, numeracy and scientific literacy

G.1 English literacy item-person map

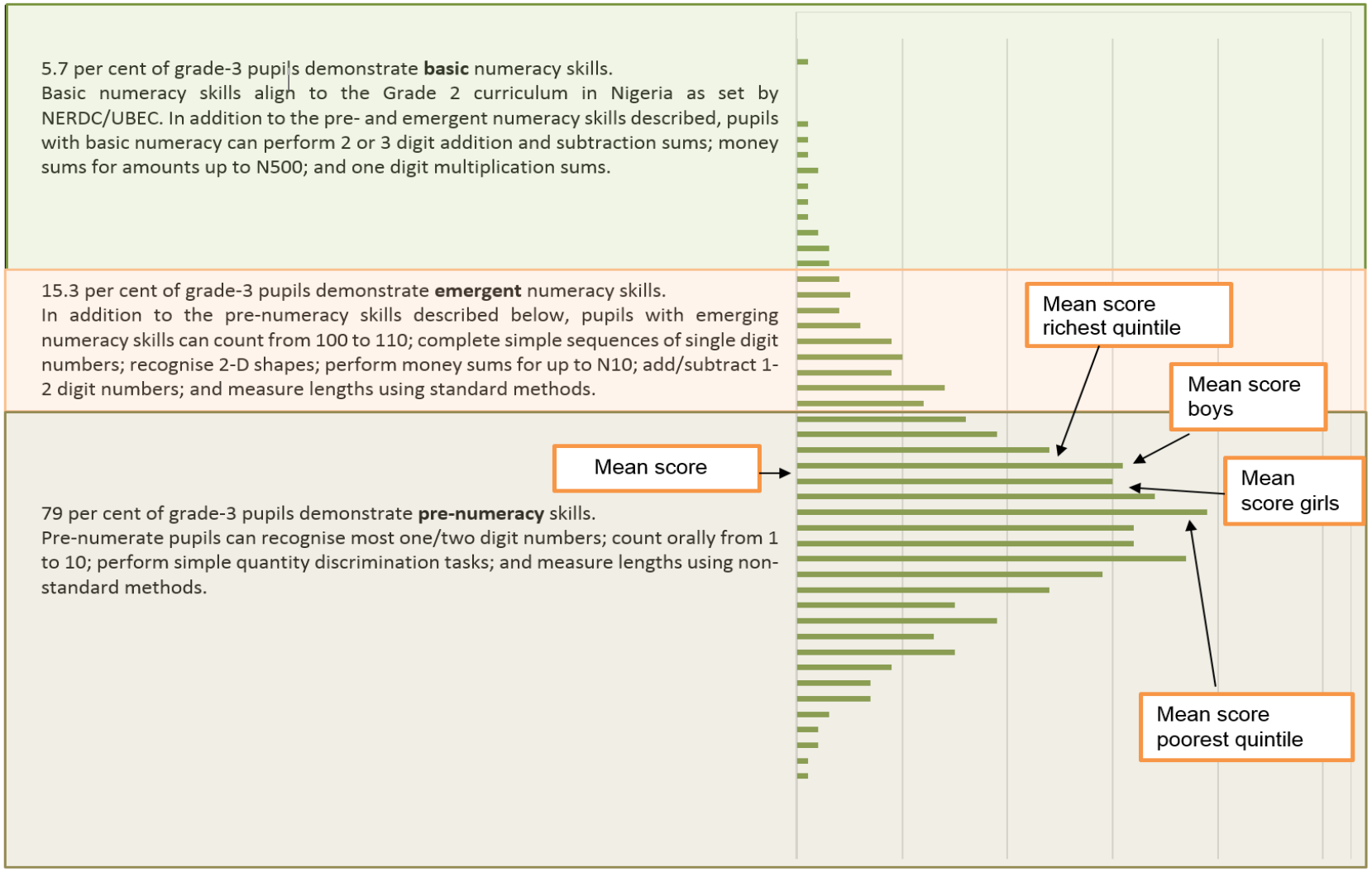
Figure 9 below captures these results along the English literacy learning trajectory, generated using Rasch analysis. These figures provide a map of the English literacy test by placing the difficulty of the items on the same measurement scale as the ability of the pupils. Items are lined up on the vertical axis, with the most difficult items (testing basic English literacy skills) at the top and the easiest items (testing pre-English literacy skills) at the bottom of the vertical axis. The green bars shows the distribution of the measured ability of the pupils from most able at the top to least able at the bottom. The pupils at the top of the map had the highest scores, while the items at the top of the map are the most difficult. The pupils at the bottom of the map earned the lowest scores, and the items at the bottom of the map are easiest. The arrows indicate the position of the typical pupil overall in the distribution ('mean score'), as well as the typical boy/girl/pupil from the richest quintile, and the typical pupil from the poorest quintile.

Figure 9 Distribution of Grade 3 pupils along the English literacy learning trajectory



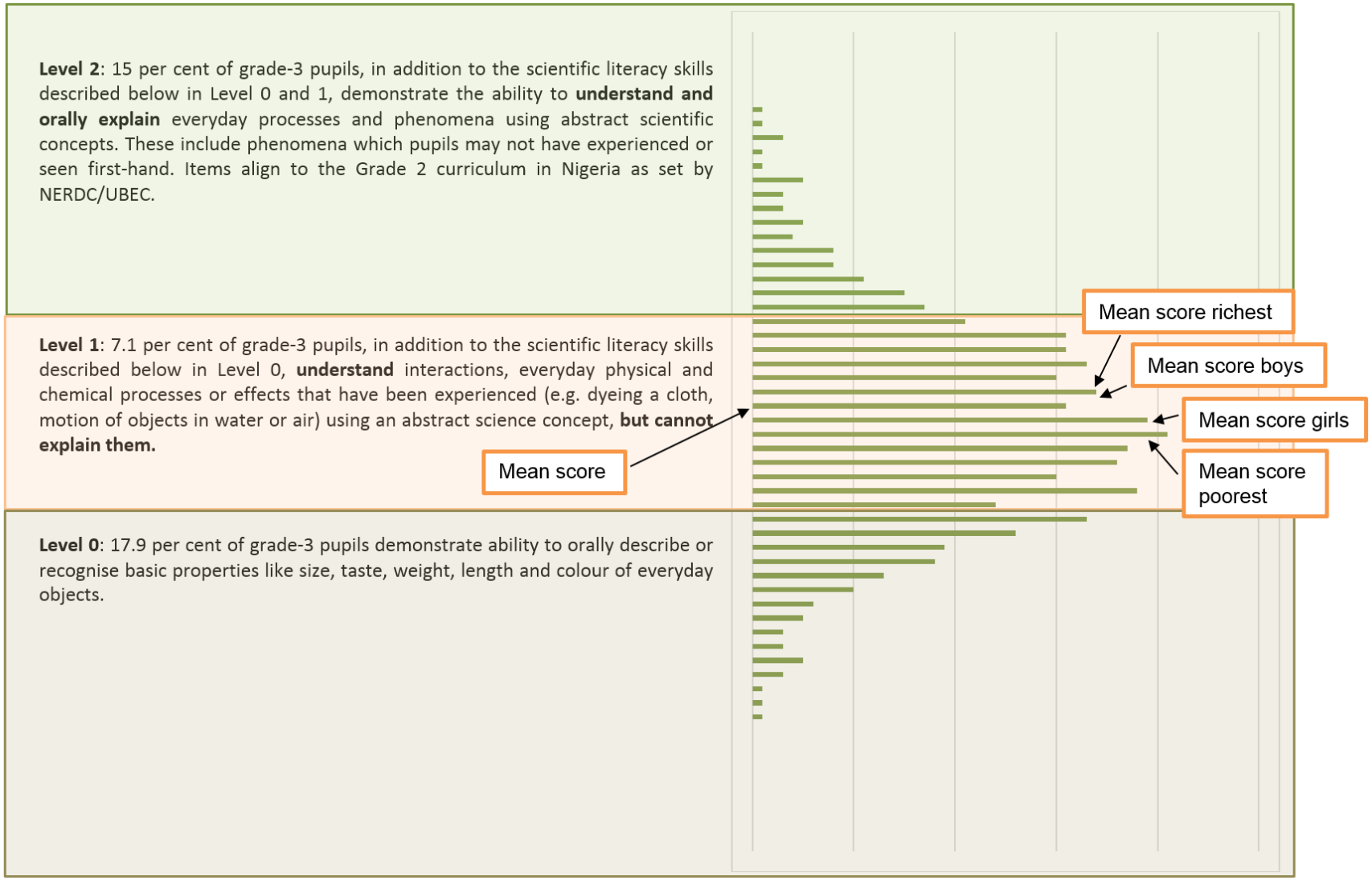
G.2 Numeracy item-person map

Figure 10 Distribution of Grade 3 pupils along the numeracy learning trajectory



G.3 Scientific literacy item-person map

Figure 11 Distribution of Grade 3 pupils along the scientific literacy learning trajectory



Annex H Pupil performance by learning domains in English literacy and numeracy: Relative strengths and weaknesses

H.1 Learning domains in literacy

In this annex, relative strengths and weaknesses in selected learning domains within English literacy (reading, writing, listening comprehension) are discussed. Apart from using Rasch modelling for analysing pupil assessment scores, raw scores were also used to examine pupil performance by various learning domains. Some key results are discussed below for reading, writing and listening comprehension skills.

H.1.1 Reading skills

H.1.1.1 Alphabet knowledge: Sounding out or naming letters from the English alphabet

To assess pre-literacy reading skills, pupils were shown a grid of 50 upper-case and lower-case letters from the English alphabet and were instructed to sound out or name as many as they could, within one minute.

As Table 22 shows, Grade 3 pupils were able to name or sound out about six letters on average. Boys named significantly more letters on average (at a 10% level) than girls. The poorest 20% of pupils named significantly fewer letters (three) compared to the richest 20% (8). Though not perfectly comparable due to different samples, instruments and grades,[39] an Early Grade Reading Assessment (EGRA) administered to Grade 3 pupils in Bauchi and Sokoto in Nigeria found even weaker results: pupils could only sound out or name three letters in Bauchi and one in Sokoto (USAID, 2013a, 2013b). There was no significant difference between pupils in treatment and control schools.

H.1.1.2 Oral reading speed: Reading familiar two-, three- and four-letter words

Knowledge of alphabets or their sounds is a fundamental building block in the acquisition of reading skills. Pupils were shown a set of 25 familiar two-, three- and four-letter words (such as: 'up', 'sad', 'good', etc.), and were instructed to read as many of them as they could in one minute. This question seeks to test emerging English literacy skills, expected by the end of Grade 1.

On average, pupils read out one word per minute. Richer pupils read significantly (at a 5% level) more words than poorer pupils, but there were no significant gender differences. Pupils in treatment schools performed significantly better than pupils in control schools (at a 10% level) – however, the effect size is small (0.02) and negligible in practice.[40]

[39] The EGRA was administered to Grade 3 pupils towards the end of their school year, while this survey tested Grade 3 pupils at the beginning of their school year. Separate EGRAs were administered for English and Hausa; the results discussed here refer to the English EGRA results only.

[40] Cohen (1988) suggested that effect size=0.2 be considered 'small', 0.5 represents a 'medium' effect size and 0.8 a 'large' effect size. This means that if two groups' means do not differ by 0.2 standard deviations or more, the difference is trivial, even if it is statistically significant.

Table 22 Reading skills: Alphabet knowledge and oral reading speed

Indicator	Overall	Treatment	Control	Boys	Girls	Poorest 20%	Richest 20%
No. of letters sounded out/named in one min.	5.5	5.3	5.8	6.0*	4.8	3**	7.6
No. of words read out in one min.	0.9	1.0*	0.8	0.9	0.9	0.7**	1.1

Source: Quantitative impact evaluation baseline survey (October 2014), pupil learning assessment. Note: (1) Statistically significant differences between groups are marked with asterisks: *significant at the 10% level; **significant at the 5% level; ***significant at the 1% level.

H.1.2 Writing skills

H.1.2.1 Copying two printed sentences following some grammar conventions

Pupils were shown two printed sentences and were instructed to copy them down in their pupil books following the grammar and punctuation conventions (for instance, comma, full stop, spacing between words, etc.) as shown in the printed text. This was deemed an emergent literacy item, testing skills expected to be acquired by the end of Grade 1. 62% of Grade 2 pupils could not copy these sentences following at least some (if not all) grammar and punctuations correctly.

H.1.2.2 Spelling: Writing at least one three- or four-letter word correctly based on oral instructions

Pupils listened to four words, which were dictated, and were instructed to write these down as a test of their ability to spell three- or four-letter words, corresponding to basic literacy skills expected by the end of Grade 2. Only about 5% of pupils could correctly spell and write down at least one of the words correctly, suggesting that for a vast majority of pupils, English writing skills in Grade 3 do not correspond to the level expected at their grade level. In both writing questions there were no significant differences between boys and girls; however, pupils from the richest 20% of households performed significantly better (at a 1% level) than pupils from the poorest 20% of households. There were no significant differences between pupils in treatment and control schools.

H.1.2.3 Listening comprehension

Pupils were read a short story (four sentences) in English and then asked two questions to assess their listening comprehension. This was deemed an emergent literacy item, testing skills expected to be acquired by the end of Grade 1. Only 13% of Grade 3 pupils answered at least one of the two questions correctly, while the rest (87%) could answer none. There were no significant differences between boys and girls, but pupils from the richest 20% of households performed significantly better (at a 1% level) than pupils from the poorest 20% of households. Again, there were no significant differences between pupils in the treatment and control schools.

H.2 Learning domains in numeracy

This section discusses relative strengths and weaknesses in selected learning domains within numeracy. Some key results for number concepts, addition and subtraction, and for other learning domains using raw scores, are discussed below.

H.2.1 Orally counting from one to 10 and from 100 to 110

Pupils were instructed to orally count from one to 10 and then from 100 to 110 (separate items). The first item was classified as testing pre-numeracy skills and the latter as testing emergent numeracy skills. About 85% of Grade 3 pupils could do the former correctly while only 25% could do the latter. In both cases, there were no significant gender differences, but there were significant differences (at a 1% level) by household wealth, with a larger proportion of pupils from richer households answering this question correctly. Pupils in control schools performed significantly (at a 5% level) better than pupils in treatment schools on the first counting question (one to 10) but the effect size is very small (0.07) and as such this difference, even though significant, is inconsequential.

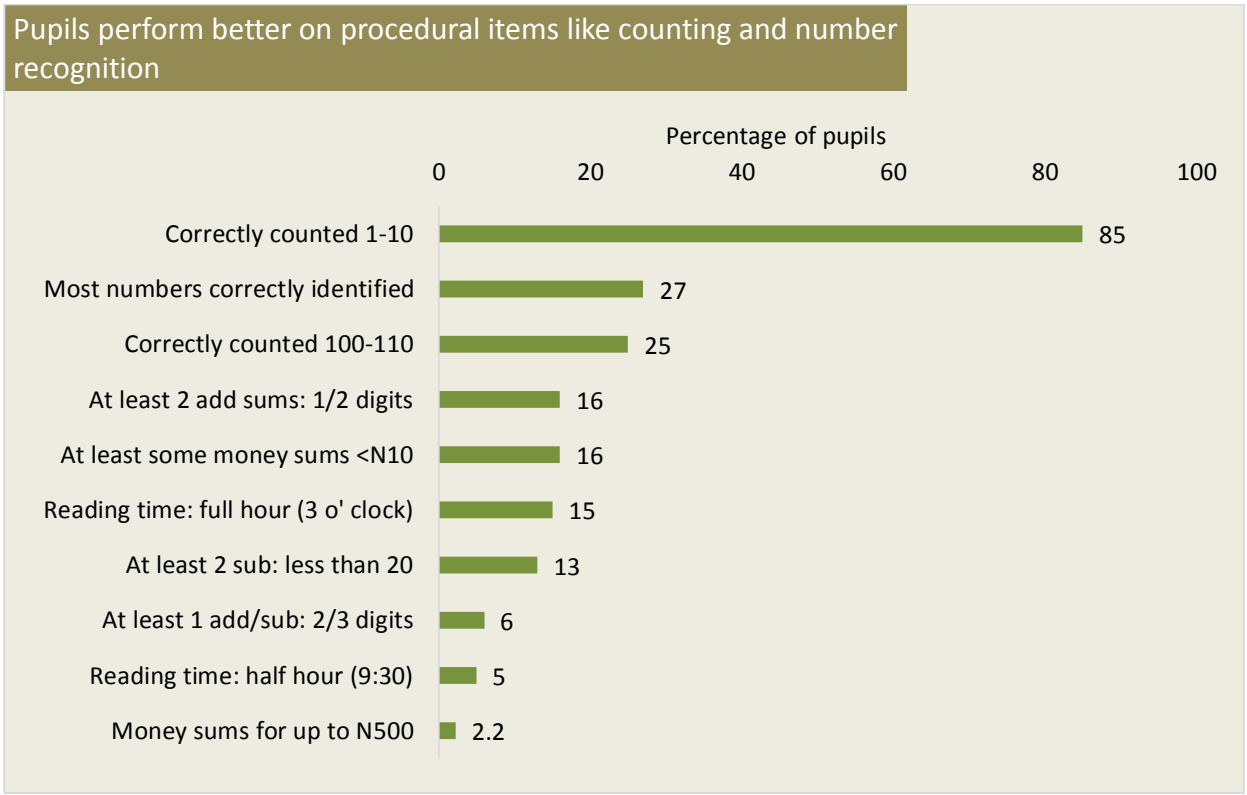
H.2.2 Orally recognising numbers

Pupils were shown a grid of 15 one- and two-digit numbers listed in no particular sequence and were instructed to orally identify them. On average, pupils identified four numbers correctly. This question was classified as testing pre-numeracy skills (level 0), and yet 40% of the Grade 3 pupils could not identify any number, indicating that they were about two levels below the numeracy skills expected at their grade. About 33% of pupils identified some numbers correctly, while 27% identified most numbers correctly (Figure 12). Boys and pupils from the richest 20% of households performed significantly better (at 1% level) than girls and pupils from the poorest 20% of households, respectively. There were no significant differences between pupils from treatment and control schools.

H.2.3 Addition/subtraction

Pupils were also asked to perform four additions with single- or double-digit numbers in four minutes. About 16% of pupils could do at least two sums correctly. Pupils performed worse on subtraction, and worse on two- and three-digit addition/subtraction questions. Richer pupils and boys performed significantly better (both at 1% level) than poorer pupils and girls, respectively. There were no significant differences between pupils in treatment and control schools.

Figure 12 Pupil performance by learning domains in numeracy



Overall, it appears that Grade 3 pupils performed better on the more procedural items (such as number identification, counting from one to 10) and less well on the conceptual items that require them to understand and apply their procedural knowledge (for instance, money sums or reading time). Pupils from the richest 20% of household consistently performed better than pupils from the poorest 20% of households. No consistent pattern of gender differences emerges even though boys performed significantly better than girls on several learning domains.

Table 23 Differences in pupil performance on numeracy learning domains by pupil characteristics

	Who performed better?		
Learning domains in numeracy	**Boys or girls?**	**Poorest or richest?**	**Treatment or control?**
Counting from one t 10	Same	Richest***	Control**
Number identification	Boys***	Richest***	Same
Counting from 100 to 110	Same	Richest***	Same
One-/two-digit addition sums	Boys***	Richest***	Same
Money sums for less than NGN 50	Boys***	Richest***	Same
One-/two-digit subtraction sums	Same	Richest***	Same
Two-/three-digit addition/subtraction sums	Same	Richest***	Same
Money sums for up to NGN 500	Same	Richest**	Same

Source: Quantitative impact evaluation baseline survey (October 2014), pupil learning assessment. Note: (1) Statistically significant differences between groups are marked with asterisks: *significant at the 10% level; **significant at the 5% level; ***significant at the 1% level.

Annex I Supplementary quantitative analysis: Teacher and head teacher background characteristics

The baseline survey found that only about 18% of the primary school teachers are female, which is largely consistent with other studies for northern Nigeria (Cameron 2015a) but is a much smaller proportion compared to the national average of 48% (UNESCO 2014). It is also noteworthy that the impact evaluation covers remote schools in which female teachers are generally less likely to be posted. Among head teachers the proportion of women is even smaller, at just 3%.

The average age for teachers is 37 years, with 12 years of teaching experience and about five years of teaching in their current school. There is notable variation in teaching experience: the 10% of teachers with least experience have been teaching for three years or less, compared to the 10% of teachers with most experience who have taught for 25 years or more (Annex M).

Table 24 Teacher and head teacher background characteristics (mean estimates)

Indicator	Overall	Treatment	Control	Male	Female	N
Teachers						
Gender (% female)	17.8	19.4	16.1	82.2	17.8	908
Age (years)	37.0	36.8	37.1	37.0	36.7	903
Experience teaching (years)	12.3	12.4	12.2	11.9***	14	900
Experience in current school (years)	5.3	5.0	5.5	5.1***	6.2	896
Holds NCE qualification (%)	67.4	68.5	66.4	64.4***	81.4	908
Received in-service training during last two years (%)	47.9	49.5	46.4	50.0***	38.2	907
Head teachers						
Gender (% female)	3.0	2.5	3.5	97.0	3.0	330
Age (years)	44.7	44.9	44.6	44.8	43.2	330
Experience teaching (years)	20.4	20.7	20.2	20.4**	23.2	329
Experience as head teacher (years)	11.0	10.7	11.3	11.2***	5.5	314
Experience as head teacher in current school (years)	3.6	3.4	3.9	3.6	4.6	318
Holds NCE qualification (%)	86.6	86.8	86.5	86.5	89.6	330
Received in-service training last two years (%)	78.8	79.0	78.6	79.7**	50.4	330
Teaches any primary class regularly (%)	66.6	60.8***	72.0	67.0	51.3	330

Source: Quantitative baseline survey (October 2014), head teacher interviews and teacher interviews. Note: (1) Statistically significant differences between groups are marked with asterisks: *significant at the 10% level; **significant at the 5% level; ***significant at the 1% level.

Head teachers are on average older (45 years), a larger proportion (87%) hold an NCE qualification, and they have more teaching experience (21 years) than teachers, as would be expected. As for teachers, there is large variation in teaching experience. The 10% of head teachers with least experience have taught for nine years or less, whereas the 10% with the most

experience have taught for 32 years or more (Annex M). The average head teacher has had this role for 11 years and has been head teacher in her/his current school for almost four years.

Just over two-thirds (67%) of teachers hold an NCE qualification,[41] consistent with the national average (NCCE 2005; EMIS, 2011), and almost half (48%) received in-service teacher training during the last two school years. Among the teachers who received in-service training, training on literacy and/or numeracy (49%) was most common, followed by training on teaching methods (43%), school leadership (10%), knowledge of curriculum subjects other than literacy and numeracy (10%), and about 1% for extra-curricular activities and school management/development planning (Annex M).[42]

Head teachers are much more likely to have received in-service training in the last two school years (79%) than teachers. The main training topics for head teachers who attended in-service training during the last two years were: literacy and/or numeracy (42%), teaching methods (30%), school leadership (19%), school management/development planning (16%), knowledge of curriculum subjects other than literacy and numeracy (7%), and development of instructional materials (2.5%) (Annex M).[43] In addition, 11% of head teachers reported 'other topics' for the training agenda, with the most common topics being: computing and information and communications technology (ICT), SBMCs, record keeping and lesson planning.

Box 7 Teacher and head teacher mobile phone ownership and usage

As part of its in-service training the TDP will use a 'trainer in the pocket' model that will provide teachers with continuous access to audio-visual resources using mobile technology. The vast majority of head teachers and teachers (91%) own a mobile phone that is in a working condition, and among these 14% have a mobile with audio and 77% have a mobile phone with both audio and video. The most common use of mobile phones is to make voice calls (96%), send text messages/SMS (85%), browse the internet (46%), use the calculator (43%), check the date and/or time (35%), watch videos (26%), listen to the radio (26%), listen to music not on the radio (24%), use social media such as Facebook (24%), and take photos or look at photos (23%). Smaller proportions of teachers also used their mobile phones for research and learning (14%), playing games (13%) and reading the news (11%).

Among the teachers, 26% teach one or more of the lower primary Grades 1–3 only, 16% teach one or more of the upper primary Grades 4–6 only, and 58% teach both the lower and upper primary grades, and there is no significant difference for teachers in the treatment and control groups (Annex M). This suggests that teaching arrangements differ across schools, with teachers in some cases being subject specialists whereas in others they are generalist classroom teachers.

A range of subjects are taught but the majority of the interviewed teachers teach: English (44%), science (43%), maths (41%), Hausa (27%) and social studies (27%). The TDP will provide training on the first three of these subjects to teachers selected to participate in the programme's in-service training. Other subjects that are taught, but by much smaller proportions of teachers, are: Islamic studies (6%), agriculture (4%), civic education (3%), drawing (3%) and writing (2%). For head teachers who teach, subjects taught largely mirror those of teachers: English (54%), maths (50%),

[41] The NCE is the minimum qualification required to teach at primary level.
[42] Teachers could report more than one training topic so the proportions of teachers trained on each topic do not sum to 100%.
[43] Head teachers could report more than one agenda topic so the proportions of head teachers trained on each topic do not sum to 100%.

science (36%), social studies (30%), Hausa (25%), agriculture (7%), Islamic studies (4%), civic education (3%), cultural and creative arts (2%) and Arabic (2%).

The proportion of head teachers who teach is significantly larger in the control group: 72%, compared to 61% in the treatment group. However, the effect size is small (0.08) and as such this difference, even though significant, is inconsequential The large proportions of head teachers that teach both in the treatment and control groups should be seen in the context of many of the schools being small (e.g. the 10% of schools that are smallest have 142 or fewer pupils), and head teachers in the smaller schools are more likely to teach than head teachers in the larger schools.

There are not only differences in background characteristics between teachers and head teachers there are also differences in regard to gender among teachers and head teachers. Although the average age is 37 years both for female and male teachers, female teachers have significantly more teaching experience (14 years compared to 12 years for men) and a significantly larger proportion of female teachers (81%) than male teachers (64%) have an NCE qualification. However, a significantly larger proportion of male teachers have attended in-service teacher training during the last two school years (50%) than female teachers (38%).

Similarly for head teachers, despite there being no significant difference in average age by gender, female head teachers have significantly more teaching experience (23 years) than male head teachers (20 years), and a significantly larger proportion of male head teachers (80%) have attended in-service training during the last two years compared to female teachers (50%). However, male head teachers on average have significantly more experience working as a head teacher (11 years) than female head teachers (six years).[44]

Box 8 Baseline results for teacher characteristics in Jigawa, Katsina and Zamfara

- The average teacher is 37–38 years old and has 12–13 years of teaching experience in the three states.
- There are major differences in the proportions of: female teachers, teachers with an NCE qualification and teachers that received in-service training across the three states:
  - the proportion of female teachers is 33% in Katsina, 12% in Zamfara and 9% in Jigawa 9%;
  - the proportion of teachers with an NCE qualification is 81% in Katsina, 61% in Zamfara and 60% in Jigawa; and
  - the proportion of teachers who received in-service training during the last two years is 74% in Jigawa, 78% in Zamfara and 30% in Katsina.
- The average head teacher in the three states is 44–46 years and has 10–12 years of experience working as a head teacher.
- The proportion of female head teachers is very small in all the states: Katsina 6%, Zamfara 2% and Jigawa 1%.
- There are large state differences in the proportions of head teachers with an NCE qualification: Katsina 92%, Zamfara 88% and Jigawa 80%.
- There are also substantial differences in the proportions of head teachers who received in-service training in the last two years: Jigawa 96%, Zamfara 78% and Katsina 62%.

For a detailed discussion of the baseline results at state level see the individual state reports (De, Pettersson, and Morris 2015; De, Pettersson, *et al*. 2015a; De, Pettersson, *et al*. 2015b).

[44] The relatively small sample size for head teachers makes statistical inference with respect to the tests for significant differences in group means or proportions less robust.

Annex J Supplementary quantitative analysis: Teachers' pedagogical skills

J.1 Results: Creating a positive classroom climate through the use of praise

The creation of a positive classroom climate, including encouraging pupils to ask questions and participate, and using praise to give positive feedback, is part of effective practice and is conducive to pupil learning. By contrast, a negative classroom climate reduces the scope for using effective teaching practices (Siraj *et al*. 2014; Westbrook 2013).

Almost 80% of teachers in the treatment and control schools used praise more frequently than reprimands, suggesting that the large majority of teachers attempt to create a positive classroom climate (Table 25). The ESSPIN composite survey, which covers six states in Nigeria including Jigawa, and uses the same indicator, also found that a similarly large proportion of teachers (80%), used praise more often than reprimands (Cameron 2015a).

Table 25 Teachers' use of praise and reprimands in the classroom

Indicator	Overall	Treatment	Control	Male	Female	N
Praise more than reprimand (% of teachers and head teachers who teach)	79.3	79.9	78.8	81.0**	75.0	1054

Source: Quantitative impact evaluation baseline survey (October 2014), classroom observation instrument. Note: (1) The above indicators include teachers as well as head teachers who teach a primary class regularly; (2) statistically significant differences between groups are marked with asterisks: *significant at the 10% level; **significant at the 5% level; ***significant at the 1% level.

However, there is a significant difference (at 5% level) across female and male teachers. The proportion of male teachers that use praise more often than reprimands was 81%, compared to 75% for female teachers. This differs from the findings of the ESSPIN composite survey conducted in six Nigerian states, which found the opposite: 87% of female teachers used praise more often than reprimands, compared to 75% of male teachers (Cameron 2015a). It appears that this difference is largely driven by female teachers from the southern states (particularly Enugu); the gender differences in this indicator are not significant in the northern states (Jigawa, Kano and Kaduna).

Box 9 The practice of co-teaching: Prevalence and characteristics

The classroom observation instrument recorded whether classrooms under observation were being co-taught or not. Co-teaching refers to the practice of two or more teachers teaching and/or managing a class. In most cases, one of the teachers 'leads' the teaching while the other teacher(s) manages the classroom, checks pupils' work, and distributes learning aids among pupils etc.

Co-teaching could take place for a number of reasons, including on-the-job training for new teachers, lack of classrooms and staffrooms, or to manage classes that are larger than might be deemed manageable for a single teacher. This survey noted occurrences of co-teaching during classroom observations but did not investigate the reasons for this.

Co-teaching introduces complications for the classroom observation instrument and data. For instance, the 'lead' teacher being observed may engage in certain activities (e.g. moving around

Box 9 The practice of co-teaching: Prevalence and characteristics

among pupils; assisting pupils) with less frequency because a co-teacher is simultaneously present in the classroom carrying out these tasks. Data collectors were trained to record the teacher talk/action of the 'lead' teacher only.

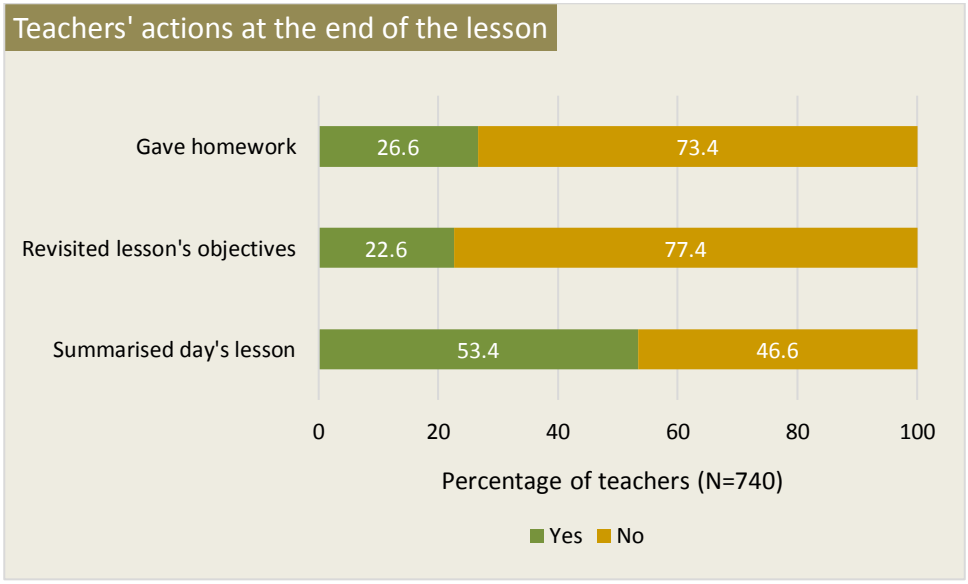
Overall, only 6% of observed classrooms were being co-taught during the observation. In lessons with co-teaching, a significantly smaller proportion of teachers used praise more than reprimands (67%) and the average class size was larger (54 pupils) compared to lessons with no co-teaching, where 80% of teachers used praise more than reprimands and the average class size was 41 pupils.

J.2 Results: Teaching practices at the end of lessons

The baseline survey also examined selected teaching practices at the end of a lesson, including summarising materials covered during the lesson, going over the objectives of the lesson to check if pupils have understood the material covered, and setting relevant homework – practices that tend to characterise effective teachers (Siraj *et al*. 2014).

Lessons were observed for up to 36 minutes as the standard lesson length is 35 minutes. Nevertheless, many lessons (37%) lasted longer than 36 minutes, meaning that data on end-of-lesson practices were not available for these lessons. At the end of the lesson, just over half of the teachers (53%) summarised their lesson; about 27% gave their pupils homework; and 23% revisited the lesson objectives (Figure 13).

Figure 13 End-of-lesson teaching practices

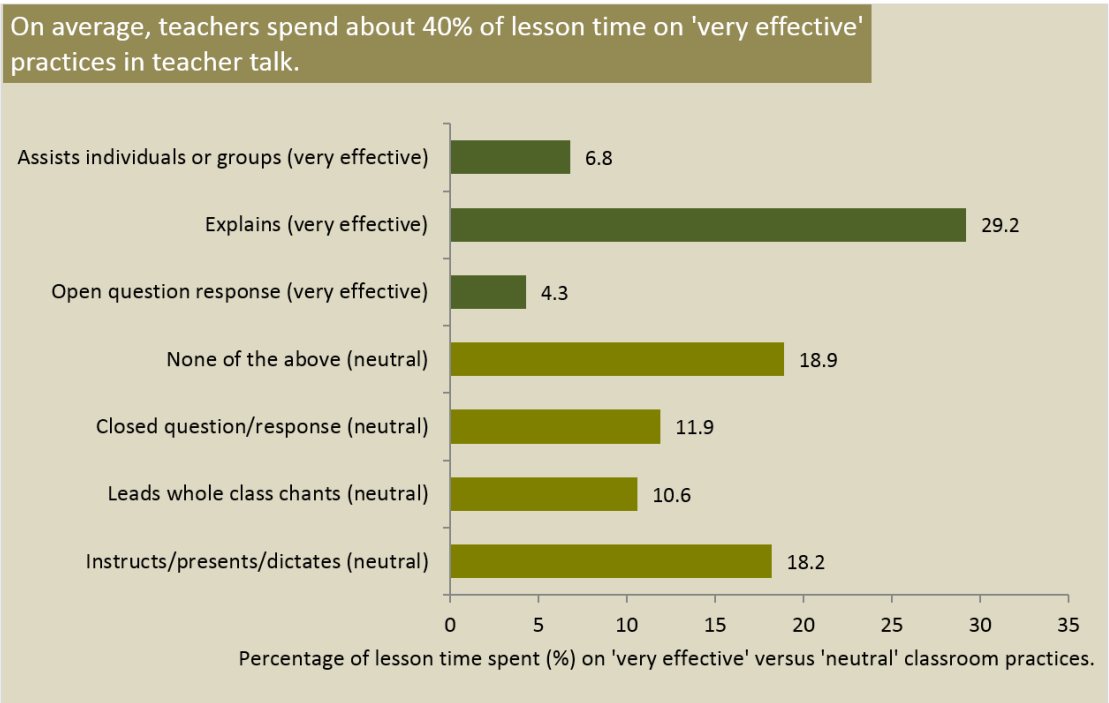


The impact evaluation baseline results are roughly similar to those of a study from Tanzania that included lesson observations from 17 districts in Tanzania. This study found that at the end of lessons, around 20% of teachers checked if pupils had grasped the new material set out in the lesson introduction or used a plenary to summarise the materials covered during the lesson (Pettersson and Rawle 2015).

To further explore teaching practices in the three states, Figure 14 provides an indication of the extent to which specific teaching practices were used during lessons. Beginning with the **teacher talk** category, on average, neutral teaching practices account for 60% of total lesson time. On the

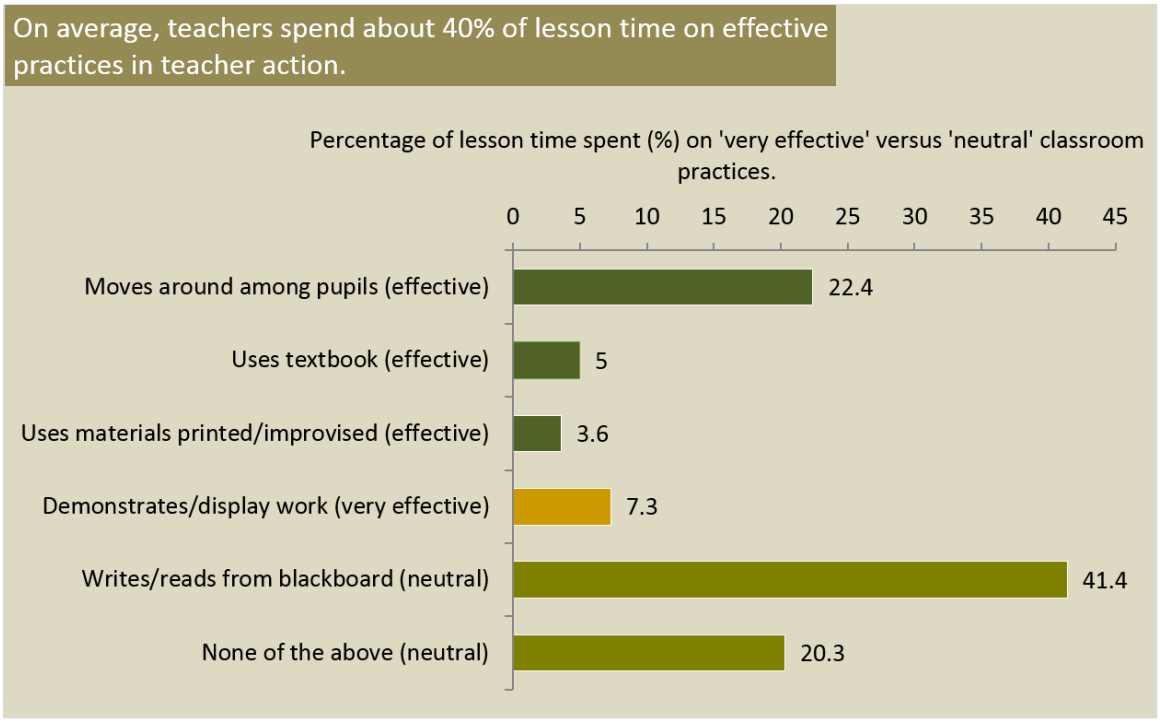
other hand, very effective practices together account for 40% of total lesson time. There were no significant differences between the sizes of various teacher talk categories for treatment and control teachers.

Figure 14 Mean distribution of lesson time by various 'teacher talk' categories (N=1054)



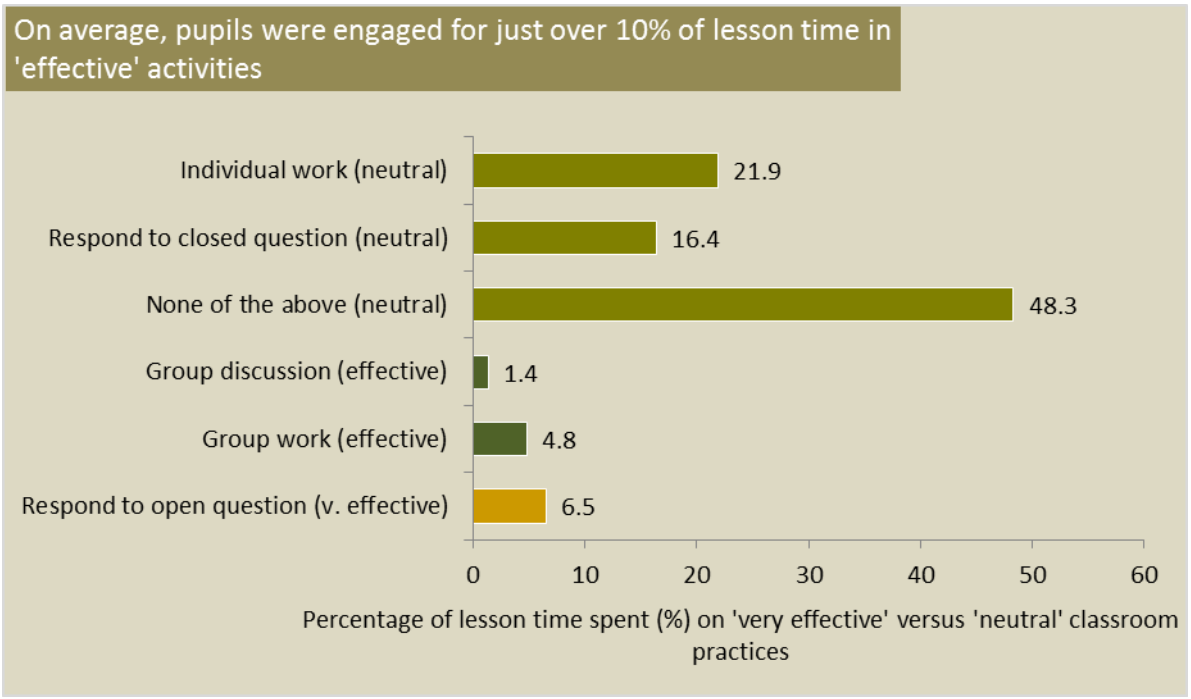
When it comes to the **teacher action** category, simply writing or reading from the blackboard, which is here classified as a neutral practice, on its own takes up a considerable 41% of total lesson time. Effective practices jointly comprise about one-third of the lesson time (31%). By contrast, it is only for 7% of the lesson time that teachers use the blackboard to demonstrate how to solve a problem or explain a concept. Similar to the teacher talk category, other practices, which may include teachers sitting at their desk reading or marking pupils' work, take up about 20% of lesson time. There were no significant differences between treatment and control teachers in these patterns of teacher action.

Figure 15 Mean distribution of lesson time by various 'teacher action' categories (N=1054)



For **pupil activities**, unspecified activities account for almost half of total lesson time (48%) and include listening to the teacher talk, waiting while the teacher writes on the blackboard and chanting, among other things. Other neutral practices combined take up 6% of lesson time. In line with the earlier result, where teachers appeared to rarely ask open questions, pupils spent a little less than 7% of the lesson time answering open questions. There were no significant differences between treatment and control teachers, except for group discussion (effective), where treatment teachers appeared to spend more time on group discussion or presentation. However, the effect size is small (0.135) and as such this difference, although significant, is inconsequential.

Figure 16 Mean distribution of lesson time by various 'pupil activity' categories (N=1054)



The above findings on teachers' limited use of effective classroom practices is supported by studies from other countries.[45] For example, a school survey of 17 districts in Tanzania found that during observed lessons the vast majority of teachers did not (1) encourage individual pupils to ask questions or explain ideas; (2) ask open-ended questions; or (3) ask pupils to carry out activities in pairs or in groups. Moreover, about half of the teachers never commented on or probed pupils' answers (Pettersson and Rawle 2015).

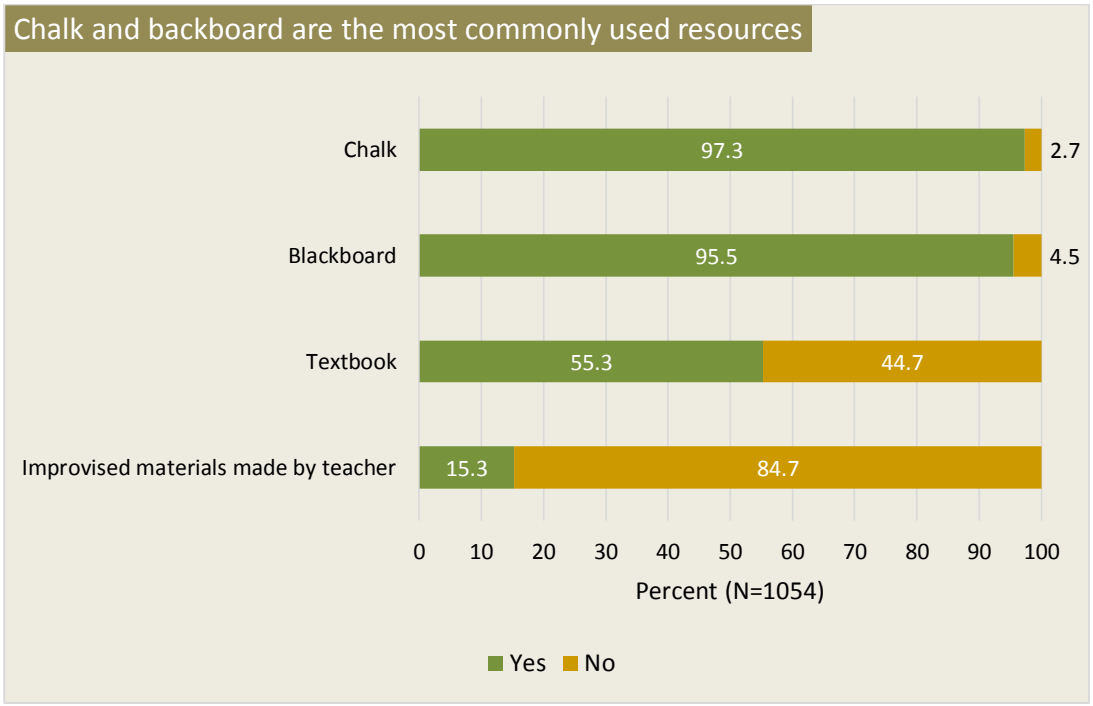
J.3 Results: Teachers' use of resources during lessons

The baseline survey also examined whether teachers used any resources during their lessons, and if so, what type. The average number of resources used was three. The most commonly used resources were chalk (97% of teachers) and chalkboard (96%), followed by textbooks (56%). Around 15% of teachers used improvised materials to aid learning; 8% used posters, charts or pictures; and 7% used resources from the local environment, such as toys, jewellery, plants and sand. There were no significant differences in patterns of resource use by teacher gender.

There were no significant differences in resources used by teachers in the treatment and control schools.

[45] All of the studies discussed here collected data on classroom practice at five-minute intervals.

Figure 17 Resources used by teachers in the classroom



Given that TDP will develop and introduce new materials for teachers and pupils (including audio materials), and will encourage teachers to use low-cost teaching aids, it is interesting to note that only 0.1% of teachers used audio equipment and no teachers used video or science equipment;[46] 15% used improvised materials. Generally, chalkboards and chalk are available in the schools while audio, science and video equipment are rare, which helps explains the pattern of resource use by teachers. The exceptions are the use of improvised materials and resources from the local environment, which are unlikely to be unaffected by a lack of classroom resources.

[46] Science lessons constituted about 15% of all the lessons observed.

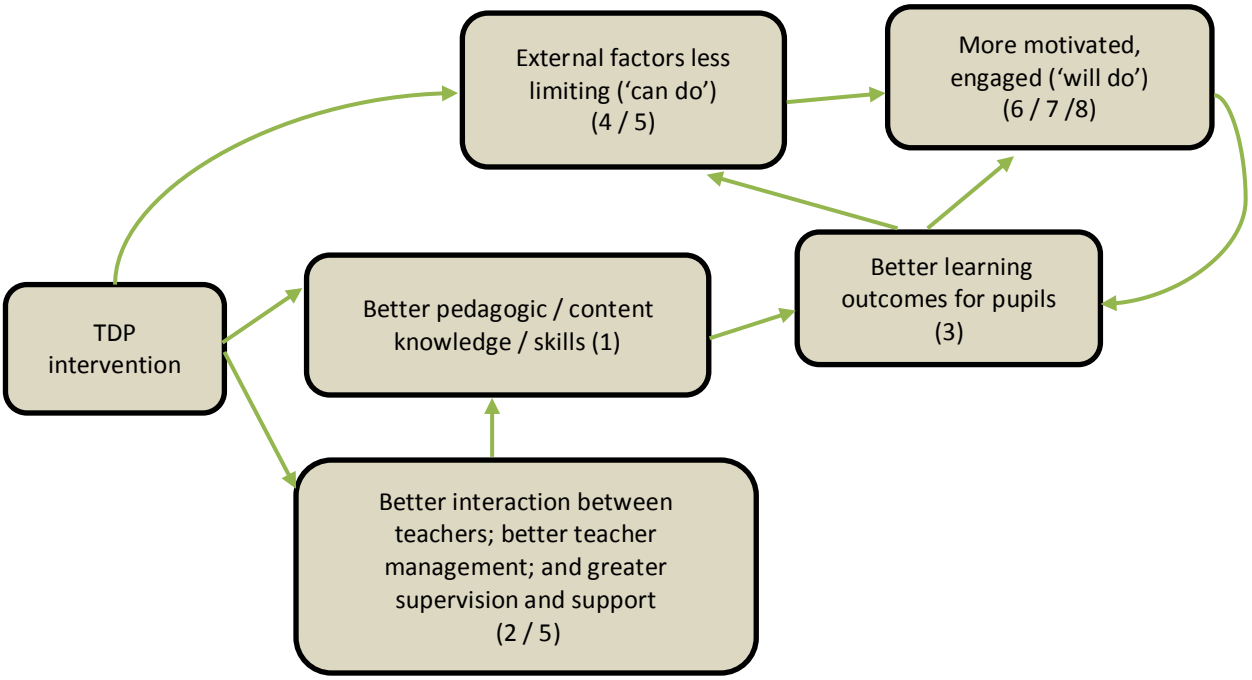
Annex K Supplementary quantitative analysis: Teacher motivation, self-efficacy, and effectiveness

For the purposes of the TDP baseline survey, teacher motivation is defined as the propensity of teachers to start and maintain behaviours that are directed towards fulfilling their professional goals, and in particular towards achieving better learning outcomes for the school's learners, based on Bennell and Akyeampong (2007) and Hoy and Miskel (1991).

The study of teacher motivation, and of work motivation more generally, often considers a wide range of factors associated with this central aspect, including conditions of work, school climate, incentives, teachers' sense of professional engagement, and occupational status. Much of the work focuses on self-efficacy: the degree to which teachers see themselves as being able to bring about the desired outcomes. It has been argued that self-efficacy is central to motivation, because teachers who believe that they cannot achieve their goals – whether they attribute this to their personal shortcomings, to aspects of the environment, or some combination of the two – are unlikely to put much effort into working towards them.

This study considers some of these associated factors, focusing in particular on the pathway suggested in the evaluation framework: 'teachers' motivation is also positively affected as they feel more "effective" and their pupils' learning outcomes improve, and this affects teachers in a number of ways, including on reduction in teacher absenteeism' (EDOREN, 2014). It is plausible to expect that a feedback mechanism could be triggered by a training intervention: seeing positive results leads teachers to reassess their own efficacy, which in turn increases their motivation, increasing their time on a task or the effort they put into planning lessons, and then leads to more positive results in terms of pupils' learning outcomes. Additionally, training could be more effective for teachers who are more motivated in the first place. The teacher motivation instrument attempts to provide enough discriminating power to disentangle some of these complex potential causal relationships and to assess whether they apply. The proposed causal chain is shown in Figure 18. The numbers in parentheses refer to sub-scales in the motivation scale described below.

Figure 18 Proposed causal chain for the study of TDP in-service training, motivation, effectiveness and pupil learning outcomes



K.1 Measuring teacher motivation

A teacher motivation scale was developed to the following specifications: (i) it had to reflect not just efficacy but several constructs related to motivation – ideally working towards an integrated model (similar to (Neves de Jesus and Lens 2005)); (ii) it had to be appropriate for the northern Nigerian social and cultural context; (iii) it had to be suitable for reading aloud in the local language (Hausa) because it could not be taken for granted that teachers would be able to read easily.

Items were selected from previous motivation instruments (see Annex A in Cameron, 2015b). The items were reviewed by an expert and piloted in the field, and minor revisions made; time constraints prevented a more comprehensive review. Exploratory factor analysis was performed to see whether the items broke down into sub-factors in the way expected. The resulting inventory has 25 items, which is enough to be delivered orally as part of a larger survey instrument, while retaining four or five items for each sub-scale.

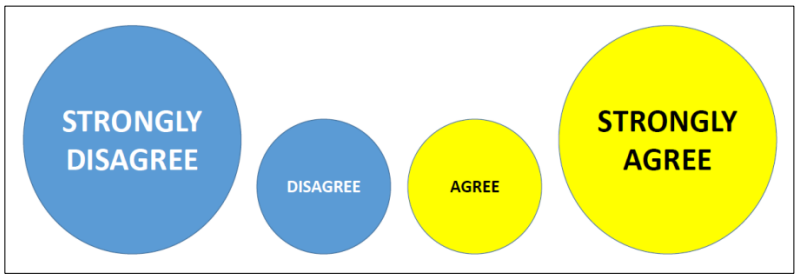
Table 26 Items in the teacher motivation inventory

Items by sub-group
Teacher–teacher interaction (2)
I have teachers that I consider my friends at my school
Teachers at my school work well together
Teachers at this school trust each other
Teachers at this school feel responsible for helping each other out
Importance of teaching effort vs. pupil background (4)
Most of the pupils in this school are not intelligent enough to do well (R)
There is no point trying to teach pupils whose parents cannot read or write (R)

Items by sub-group
I try my best to teach my pupils but their parents do not help (R)
Pupils work hard in this school and want to succeed
Importance of teaching effort vs. other circumstances (5)
I have all the support I need to teach my pupils well
Teachers at my school have the knowledge and skills to do their jobs well
It is difficult to teach in this school because the building is in a poor condition (R)
I don't always have the materials I need to do my job (R)
I have the freedom to try new things in the classroom. [Removed during the review and field pilots.]
Effort / importance (6)
Teaching well is important to me
Teachers at this school are highly committed to their job
There is no point in spending a lot of time preparing for a class (R)
It is worth working harder to make sure the pupils do well
As a teacher, I perform an important role in society
Interest / enjoyment (7)
In the past two years, my job has become more satisfying
I enjoy teaching very much
If I could choose another job today, I would still choose teaching
Teaching my class today/yesterday was boring (R)
Pressure / tension (R) (8)
Teaching is very tiring
I teach too many classes
It is difficult to manage pupils in my classrooms
There are too many pupils in my classroom
(R) indicates a reverse indicator or set of indicators, i.e. teachers who disagree with the statement are more motivated.

The order of the items was randomised but kept consistent across respondents. Items were read out aloud in Hausa and teachers indicated their agreement using a simple visual scale, shown below in Figure 19.

Figure 19 Prompt for teachers to signal agreement with statements in the teacher motivation scale



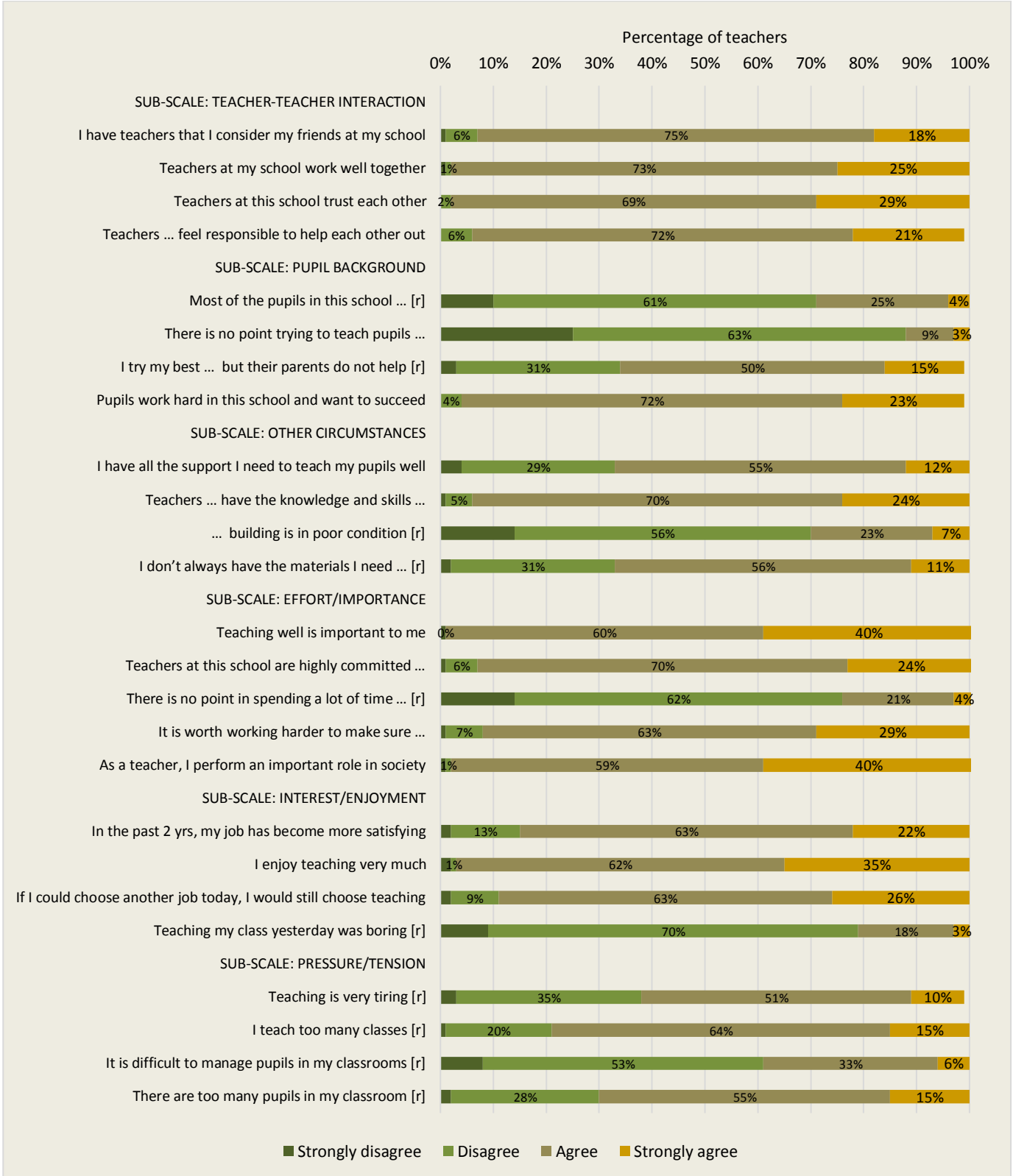
A single dimensional model for overall motivation was fitted to the data. On the basis of exploratory analysis some items were re-scored and three were dropped because of poor fit to the overall model. These were: 'I teach too many classes'; 'There are too many pupils in my

classroom'; and 'I don't always have the materials I need to do my job well'. The resulting model has good reliability: Person Separation Reliability of 0.80 and Cronbach's alpha of 0.80 (see Cameron, 2015b for further details).[47]

Results showed that a large majority of teachers agree with positively-worded statements designed as indicators of engagement and teacher–teacher interaction (Figure 20). Some 89% of teachers agreed that if they could choose another job today they would choose teaching, and 94% agreed that teachers at their school were highly committed. The results from individual items are subject to likely response biases, and teachers may be reluctant to express negative views about their own schools. The aim of the motivation scale is not to understand exactly what teachers really think, but to construct an overall scale (or set of scales) that can be used to measure the motivation of a group of teachers relative to another group, or relative to the same group at a different time period. The scale has no meaning in absolute terms, but is used for making comparisons across groups or over time.

[47] After some examination of factor loadings, a four-dimensional model was proposed, with items in slightly different categories from those proposed at the outset. However, the dimensions continue to have low reliability even though the overall scale seems be reliable, as reflected in the Cronbach alpha. Factors did not load on the items in the ways expected, and it was not possible to fit a multidimensional model of motivation to the data. It is not possible to say *a priori* whether this is a problem with the conceptual model or with the items chosen to measure it. More work is needed in regard to conceptualising and measuring the different aspects of motivation and determinants of motivation in this type of context. It would be useful to pilot a larger number of items from different scales, perhaps including task-specific motivation measures. However, the analysis does have a reliable overall measure of motivation based on the set of items as a whole.

Figure 20 Teachers' responses to items in the motivation scale[48]



[48] Teachers also had the option of refusing to respond to a statement, and in this case responses will not add up to 100% in Figure 20.

Annex L Supplementary quantitative analysis: Characteristics of 'effective' teachers

L.1 There exists a small group of relatively more effective teachers

This survey has identified four key constraints on teachers' effectiveness in the treatment and control schools at baseline: weaknesses in teachers' subject knowledge, limited use of effective teaching practices, teacher absenteeism and loss of instructional time. This section examines more closely the knowledge, behaviours and characteristics of TDP treatment and control teachers (and head teachers who teach) to understand if there are some groups of teachers who do relatively better on the aspects of teacher effectiveness captured by this baseline survey.

This section divides teachers and head teachers who teach into four groups of teachers with sufficient, near-sufficient, emerging and limited subject knowledge in maths (see Table 6 for maths achievement level definitions) to examine linkages between different aspects of teacher effectiveness (subject knowledge, pedagogy and absenteeism) and teacher characteristics. (The results are similar in terms of rankings across the four maths achievement level groups if teachers are instead grouped by their level of achievement in English subject knowledge.)[49]

L.2 Results: Teacher subject knowledge and behaviour by maths achievement level

Teacher subject knowledge is generally very weak. Figure 21 shows that average subject knowledge in English and science is significantly different (at a 5% level) across the four groups of teachers (using the 'sufficient' maths subject knowledge group as the reference point). The average English and science scores are 38% and 36%, respectively, for teachers with sufficient maths subject knowledge, but this declines to 14% for English and science for teachers with only limited maths subject knowledge. This means that teachers who do relatively well on one subject also does relatively well on the remaining two subjects as well, and vice versa.

Teacher ability to assess and monitor pupil progress is also very weak, and the difference is statistically significant (at a 5% level of significance) for teachers with different levels of maths subject knowledge (Figure 21). The average score for teachers with sufficient maths subject knowledge is 35%, compared to 20% for teachers with near-sufficient maths subject knowledge, 11% for teachers with emerging maths subject knowledge and 5% for teachers with limited maths subject knowledge. This indicates that teachers who do better in terms of maths subject knowledge also do better in terms of knowledge to assess and monitor pupils' academic progress, but that there is considerable scope for improvement for all teachers.

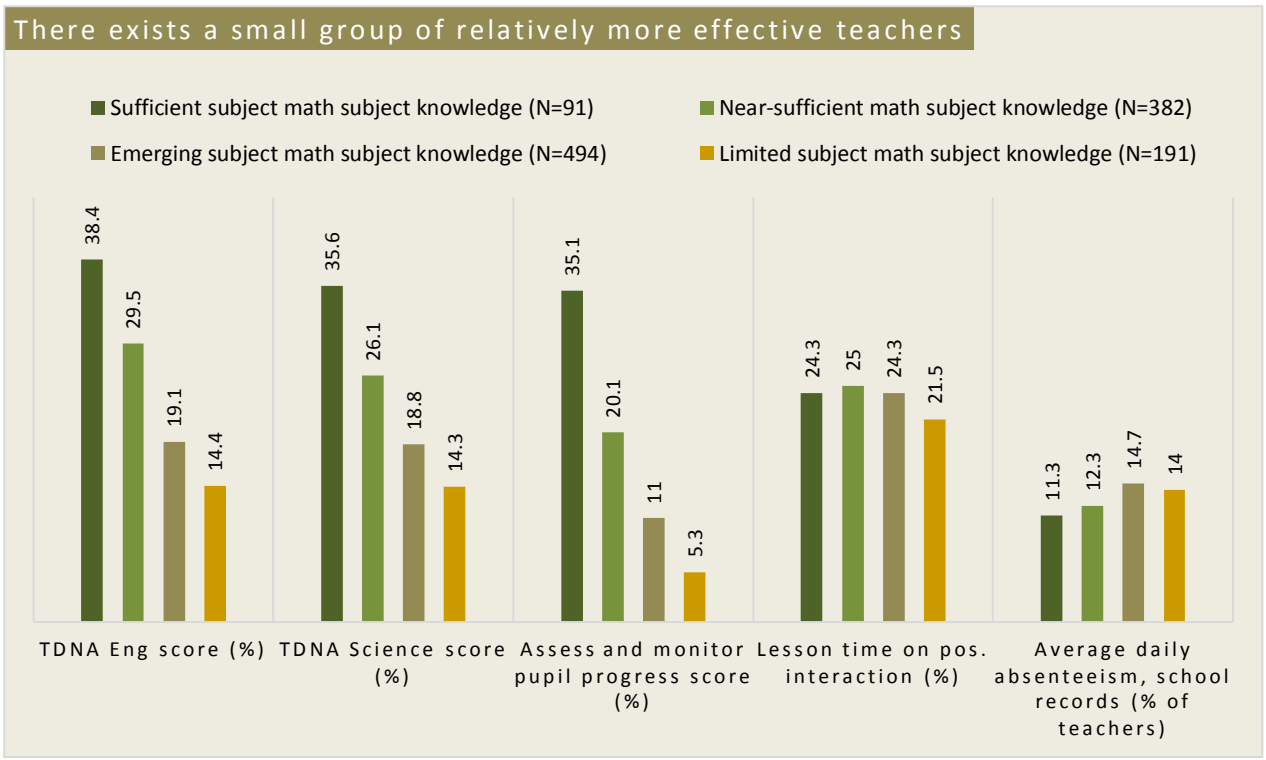
By contrast, teachers in all four groups spend a similarly (small) proportion of total lesson time (22%–25%) in positive interaction with pupils, strongly suggesting that all teachers, irrespective of their achievement level in maths subject knowledge, would benefit from in-service training in pedagogy (Figure 21).

[49] For clarity and to allow for testing of significant differences across the groups, only the results by maths achievement level are presented here. When teachers are grouped by English subject knowledge achievement level only 0.35% of teachers (four teachers) are in the group with sufficient subject knowledge, which means that statistical inference is not possible.

Average daily absenteeism is the lowest for teachers with sufficient maths subject knowledge (11%), and rises to 12% for teachers with near-sufficient knowledge (significant at a 10% level), 15% for teachers with emerging knowledge (significant at 1%), and is 14% for the group of teachers with limited maths subject knowledge (significant at 5%). The difference between 'emerging' (14.7%) and 'limited' (14%) subject knowledge groups is not statistically significant.

Overall, the available evidence suggests that levels of subject knowledge and knowledge of assessment and monitoring of pupil progress are very low. However, there exists a small group (8%) of relatively more effective teachers who on average have relatively stronger (albeit still weak) maths, English and science subject knowledge and ability to assess and monitor pupil progress, and somewhat lower absenteeism. However, this relatively more effective group of teachers still displays limited use of effective teaching practices, similar to the other three groups of teachers.

Figure 21 Comparison of teacher knowledge and behaviours by maths subject knowledge achievement level



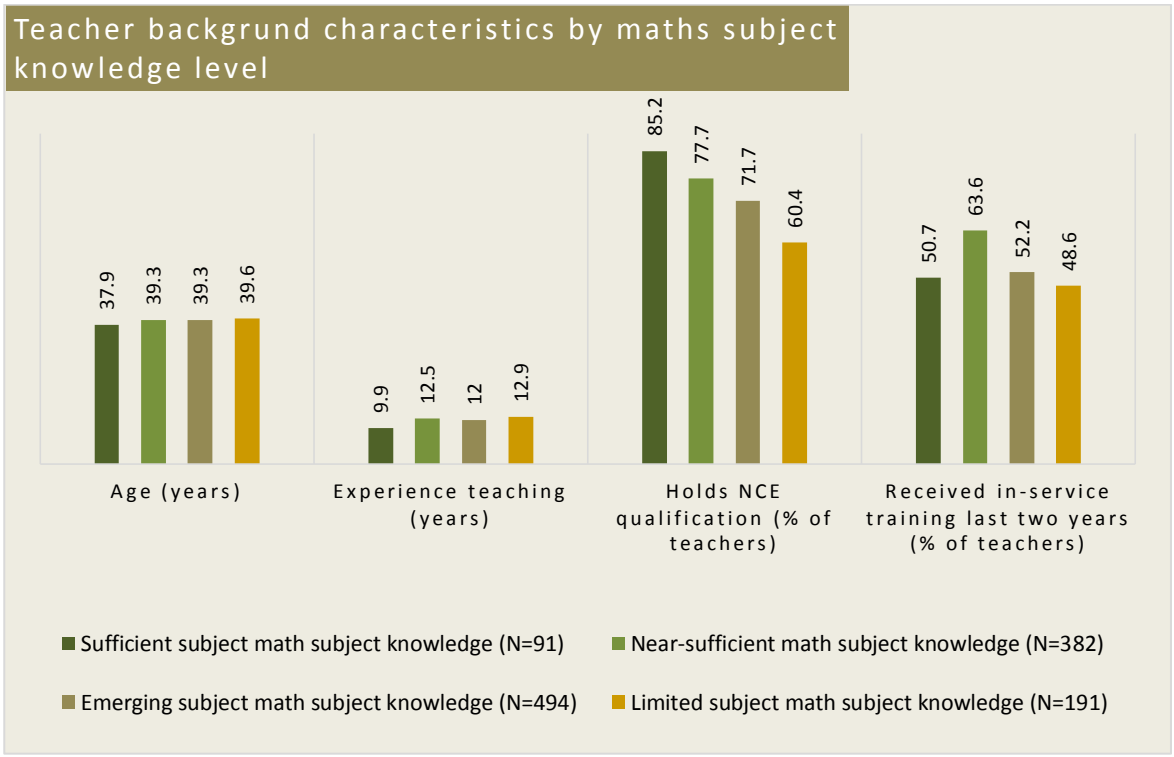
L.3 Results: Teacher characteristics by maths subject knowledge achievement level

The average teacher age ranges from 38 to 40 years across the four achievement level groups, and years of teaching experience only varies slightly, i.e. teachers with sufficient maths subject knowledge on average have significantly less experience (10 years) than teachers in the 'emerging' and 'limited' groups but not the 'near-sufficient' group (Figure 22).

The proportion of teachers who hold an NCE is larger among teachers with sufficient maths subject knowledge (85%) than among teachers with near-sufficient subject knowledge (78%) and limited subject knowledge (72%), and is much larger than for the group with only limited maths subject knowledge (60%). Taking the group of relatively more effective teachers as the reference group, these differences are significantly different (at a 5% level).

When it comes to in-service training, there are also some differences across the groups but these are not statistically significant. 53% of teachers with sufficient maths subject knowledge have received training during the last two years, 64% in the group with near-sufficient knowledge, 51% among teachers with emerging subject knowledge and 49% of teachers with limited subject knowledge.

Figure 22 Comparison of teacher characteristics by maths subject knowledge achievement level



There are no studies that provide direct comparisons for these results. However, the TDNA studies from Jigawa and Katsina discussed earlier (Johnson, D. 2008; Johnson and Hsieh 2014) provide complementary information. The study in Katsina in 2014 found that teachers who were 20–40 years old and had six to 10 years of teaching experience had the highest total scores on a TDNA that covered teaching primary maths and English, assessment of pupil writing and feedback, monitoring of pupil progress and academic literacy. Moreover, teachers with an NCE qualification scored higher (23%) than teachers with no qualifications (12%), Grade 2 (15%) or an Ordinary National Diploma (OND) (18%), but lower than teachers with a Higher National Diploma (HND) (27%), a BA in education (29%) or a BA in another subject (33%) (Johnson and Hsieh, 2014). The 2010 Jigawa TDNA was similar in content to that administered in Katsina. It found no differences in TDNA score by teacher age, but teachers with 11–20 years of teaching experience scored significantly higher than other groups, and those with 20 or more years of experience significantly lower. Teachers with an NCE scored somewhat higher (20%) than teachers with Grade 2 (18%) and OND (21%), but slightly lower than teachers with an HND (22%), and significantly lower than teachers with a degree (Johnson, 2010).

Annex M Supplementary quantitative analysis: Detailed statistical tables of baseline results

Table 27 Pupil background characteristics: Mean estimates

Indicator	Overall	Treatment	Control	Male	Female	N
Sex distribution of pupils (% of female pupils)	41.5	40.2	42.8	58.5	41.5	2575
Age in years	9.0	9.0	9.1	9.1	8.9	1908
Appropriate age (%): 8–9 yrs.	45.4	47.7	43.1	46.1	44.5	849
Underage (%): less than 8 yrs.	17.8	18.7	17	16.1*	20.2	384
Overage (%): 10 yrs. or older	36.7	33.6*	39.9	37.8	35.3	675
Speak Hausa at home (%)	98.6	98.4	98.8	98.9	98.2	2555
Mean value of household asset index	0.556	0.55	0.56	0.4***	0.78	2555

Source: Quantitative baseline survey (October 2014), Grade 2 pupil tests. Note: (1) Statistically significant differences between groups are marked with asterisks: *significant at the 10% level; **significant at the 5% level; ***significant at the 1% level.

Table 28 Pupil learning levels in English literacy: Mean estimates

Indicator	Overall	Treatment	Control	Male	Female	Poorest 20%	Richest 20%	N
Literacy scaled score	500	501.3	498.6	504.0*	494.3	464***	523	2571
Pre-literacy (%)	60.8	62.3	59.3	60.5	61.2	79.9***	49.8	2571
Emergent literacy (%)	35.8	34.7	36.9	35.3	36.5	19.7***	45.2	2571
Basic literacy (%)	3.4	3	3.8	4.2*	2.3	0.4***	5	2571

Source: Quantitative baseline survey (October 2014), Grade 2 pupil tests. Note: (1) Statistically significant differences between groups are marked with asterisks: *significant at the 10% level; **significant at the 5% level; ***significant at the 1% level.

Table 29 Pupil learning levels in numeracy: Mean estimates

Indicator	Overall	Treatment	Control	Male	Female	Poorest 20%	Richest 20%	N
Numeracy scaled score	500	494.7	505.4	506.8***	490.4	473.5	506.7	2571
Pre-numeracy (%)	79	79.8	78.1	75.4***	84	90.2***	69.9	2571
Emergent numeracy (%)	15.3	15.4	15.3	18.1***	11.5	9.1***	22.5	2571
Basic numeracy (%)	5.7	4.8	6.6	6.6	4.5	0.6***	7.6	2571

Source: Quantitative baseline survey (October 2014), Grade 2 pupil tests. Note: (1) Statistically significant differences between groups are marked with asterisks: *significant at the 10% level; **significant at the 5% level; ***significant at the 1% level.

Table 30 Pupil learning levels in scientific literacy: Mean estimates

Indicator	Overall	Treatment	Control	Male	Female	Poorest 20%	Richest 20%	N
Scientific literacy scaled score	500	496.4	503.6	505.7**	491.9	473.5***	506.7	2571
Level 0: Observes (%)	17.9	18.9	16.8	16.2*	20.3	23.5	18.7	2571
Level 1: Understands (%)	67.1	67.1	67.1	66.8	67.6	67	65.4	2571
Level 2: Explains with understanding (%)	15	14	16.1	17.0**	12.2	9.6*	15.9	2571

Source: Quantitative baseline survey (October 2014), Grade 2 pupil tests. Note: (1) Statistically significant differences between groups are marked with asterisks: *significant at the 10% level; **significant at the 5% level; ***significant at the 1% level.

Table 31 Supplementary analysis of pupil learning levels: Percentiles, confidence intervals, and standard errors

Indicator	Mean estimate	P10	P90	Standard error	Lower 95% CI	Upper 95% CI	N (overall)
Age (years)	9	7	12	0.1	8.9	9.2	1908
Literacy scaled score	500	364.1	607.1	4.4	491.3	508.7	2571
Numeracy scaled score	500	377.9	632.8	4.9	490.3	509.7	2571
Science scaled score	500	388.5	614.2	4.9	490.3	509.7	2571
Literacy raw score (out of 100)	17.2	3.1	35.9	0.8	15.6	18.7	2575
Numeracy raw score (out of 100)	27.4	11.8	54.4	0.9	25.5	29.2	2575
Science raw score (out of 100)	45.8	12.1	80.6	1.2	43.5	48.1	2575

Source: Quantitative baseline survey (October 2014), Grade 2 pupil tests. Note: (1) Statistically significant differences between groups are marked with asterisks: *significant at the 10% level; **significant at the 5% level; ***significant at the 1% level.

Table 32 Head teacher and teacher ownership and use of mobile phones

Indicator	Overall	Treatment	Control	N
Own a working mobile phone (% of head teachers and teachers)	90.5	89.7	91.3	1103
Type of mobile phone owned (% of head teachers and teachers who own a mobile phone)				
Audio	13.9	12.3	15.6	170
Audio and video	76.6	77.4	75.7	933
Uses mobile phone to (% of head teachers and teachers who own a mobile phone)				
Voice calls	95.8	95.4	96.1	1103
Send/receive SMS (text) / MMS	84.4	84.6	84.3	1103
Listen to radio	26.0	24.6	27.5	1103
Listen to music (non-radio)	23.8	22.2	25.5	1103
Play games	13.1	12.2	14.1	1103
Take photos/look at photos	23.0	21.8	24.1	1103
Browse internet	45.7	42.6**	48.8	1103
Read news/stories	11.4	10.6	12.3	1103
Calculator	43.1	42.7	43.5	1103
Date and/or time	35.0	33.7	36.4	1103
Social media (e.g. Facebook)	23.8	22.1	25.6	1103
Watch video	26.1	24.6	27.5	1103
Research/learning	13.7	12.9	14.5	1103
Other	2.6	3.6***	1.6	1103

Source: Quantitative baseline survey (October 2014), head teacher interviews and teacher interviews. Note: (1) The categories for mobile phone use are not exclusive; (2) statistically significant differences between groups are marked with asterisks: *significant at the 10% level; **significant at the 5% level; ***significant at the 1% level.

Table 33 Grades taught by teachers

Indicator	Overall	Treatment	Control	N
Teacher teaches one or more lower primary Grades 1–3 (% of teachers)	26.1	25.5	26.7	1044
Teacher teaches one or more upper primary Grades 4–6 (% of teachers)	16.0	14.4	17.6	1044
Teacher teaches both the lower and upper primary grades (% of teachers)	57.9	60.0	55.7	1044

Source: Quantitative baseline survey (October 2014), TDNA. Note: (1) Statistically significant differences between groups are marked with asterisks: *significant at the 10% level; **significant at the 5% level; ***significant at the 1% level.

Table 34 Subjects taught by teachers and head teachers

Indicator	Overall	Treatment	Control	N
Teachers: Subjects taught (% of teachers)				
English	43.9	44.1	43.7	907
Maths	41.0	41.1	40.8	907
Social studies	26.5	25.5	27.5	907
Science and technology	43.1	44.0	42.0	907
Drawing	3.2	3.3	3.1	907
Writing	1.7	1.4	2.0	907
Islamic studies	5.8	6.6	4.9	907
Christian religious	0.2	0.3	0.0	907
Home economics	0.6	1.2***	0.0	907
Agriculture	4.3	3.3*	5.3	907
Civic education	3.2	3.4	2.9	907
Hausa	26.8	27.5	26.2	907
Other	8.1	9.3*	6.7	907
Head teachers: Subjects taught (% of head teachers)				
English	53.9	57.5	49.4	218
Maths	49.3	53.4	44.2	218
Social studies	30.1	30.7	29.4	218
Science and technology	35.8	35.3	36.3	218
Islamic studies	3.9	4.5	3.1	218
Christian religious studies	0.0	0.0	0.0	218
Home economics	0.0	0.0	0.0	218
Agriculture	7.3	6.6	8.3	218
Civic education	3.2	3.3	3.0	218
Hausa	24.5	25.8	22.7	218
Arabic	1.8	1.7	2.0	218
Cultural and creative arts	2.3	1.6	3.2	218
Other	1.9	0.0***	4.3	218

Source: Quantitative baseline survey (October 2014), head teacher interviews and teacher interviews. Note: (1) Statistically significant differences between groups are marked with asterisks: *significant at the 10% level; **significant at the 5% level; ***significant at the 1% level.

Table 35 Supplementary analysis of teacher and head teacher background characteristics

Indicator	Mean	P10	P90	Standard error	Lower 95% confidence interval (CI)	Upper 95% CI	N
Teachers							
Gender (% female)	17.8	N.A.	N.A.	1.1	15.6	19.9	908
Age (years)	37.0	27.0	49.0	0.2	36.5	37.4	903
Experience teaching (years)	12.3	3.0	25.0	0.2	11.9	12.7	900
Experience in current school (years)	5.3	0.0	12.0	0.2	5.0	5.6	896
Holds NCE qualification (%)	67.4	N.A.	N.A.	1.3	64.8	70.1	908
Received in-service training during last two years (%)	47.9	N.A.	N.A.	1.3	45.3	50.5	907
Head teachers							
Gender (% female)	3.0	N.A.	N.A.	0.8	1.5	4.5	330
Age (years)	44.7	34	54	0.3	44.1	45.4	330
Experience teaching (years)	20.4	9	32	0.3	19.8	21.1	329
Experience as head teacher (years)	11	1	26	0.4	10.2	11.7	314
Experience as head teacher in current school (years)	3.6	0	9	0.2	3.2	4	318
Holds NCE qualification (%)	86.6	N.A.	N.A.	1.5	83.7	89.6	330
Received in-service training during last two years (%)	78.8	N.A.	N.A.	1.6	75.7	81.9	330
Teaches any primary class regularly (%)	66.6	N.A.	N.A.	2.1	62.5	70.7	330

Source: Quantitative baseline survey (October 2014), head teacher interviews and teacher interviews. Note: (1) Statistically significant differences between groups are marked with asterisks: *significant at the 10% level; **significant at the 5% level; ***significant at the 1% level.

Table 36 Teachers' subject knowledge of Grade 4 English: Mean estimates

Indicator (% of teachers in…)	Overall	Treatment	Control	Male	Female	N
Level 1: Sufficient professional knowledge	0.4	0.4	0.4	0.4**	0	1158
Level 2: Near-sufficient professional knowledge	4.8	3.5**	6.1	4.7	4.9	1158
Level 3: Emerging professional knowledge	41.6	43.6	39.4	42.6*	35.2	1158
Level 4: Limited professional knowledge	53.3	52.5	54.1	52.2*	59.9	1158

Source: Quantitative baseline survey (October 2014), TDNA. Note: (1) Statistically significant differences between groups are marked with asterisks: *significant at the 10% level; **significant at the 5% level; ***significant at the 1% level.

Table 37 Teachers' subject knowledge of Grade 4 mathematics: Mean estimates

Indicator (% of teachers in…)	Overall	Treatment	Control	Male	Female	N
Level 1: Sufficient professional knowledge	7.8	6.5*	9.1	8.2*	5.3	1158
Level 2: Near-sufficient professional knowledge	32.6	31.8	33.4	32	35.5	1158
Level 3: Emerging professional knowledge	42.7	46.2***	39.1	43.6**	36.7	1158
Level 4: Limited professional knowledge	17	15.6	18.4	16.2**	22.6	1158

Source: Quantitative baseline survey (October 2014), TDNA. Note: (1) Statistically significant differences between groups are marked with asterisks: *significant at the 10% level; **significant at the 5% level; ***significant at the 1% level.

Table 38 Teachers' subject knowledge of Grade 4 science and technology: Mean estimates

Indicator (% of teachers in…)	Overall	Treatment	Control	Male	Female	N
Level 1: Sufficient professional knowledge	0.1	0	0.2	0.1	0	1158
Level 2: Near-sufficient professional knowledge	4.3	4.6	4.1	4.1	6.1	1158
Level 3: Emerging professional knowledge	32.7	31.6	33.8	32	36.2	1158
Level 4: Limited professional knowledge	62.9	63.8	61.9	63.8*	57.6	1158

Source: Quantitative baseline survey (October 2014), TDNA. Note: (1) Statistically significant differences between groups are marked with asterisks: *significant at the 10% level; **significant at the 5% level; ***significant at the 1% level.

Table 39 Teachers' ability to assess and monitor pupils' academic progress: Mean estimates

Indicator (% of teachers in…)	Overall	Treatment	Control	Male	Female	N
Level 1: Sufficient professional knowledge	0.3	0.3	0.3	0.2	0.4	1158
Level 2: Near-sufficient professional knowledge	4.8	5.6	4	5.3***	1.6	1158
Level 3: Emerging professional knowledge	14.7	14.1	15.5	14.7	15.5	1158
Level 4: Limited professional knowledge	80.2	80.1	80.3	79.9	82.5	1158

Source: Quantitative baseline survey (October 2014), TDNA. Note: (1) Statistically significant differences between groups are marked with asterisks: *significant at the 10% level; **significant at the 5% level; ***significant at the 1% level.

Table 40 Supplementary analysis of teachers' subject knowledge and pedagogical practices: Percentiles, confidence intervals, and standard errors

Indicator	Mean	P10	P90	Standard Error	Lower 95% CI	Upper 95% CI	N
English	23.1	0.0	43.5	0.4	22.3	23.9	1158
Maths	45.1	20.0	72.0	0.5	44.0	46.1	1158
Science and technology	21.7	4.5	40.9	0.4	20.9	22.4	1158
Teacher's ability to assess pupils' academic progress	14.7	0.0	40	0.5	13.8	15.7	1158
Proportion of lesson time spent in positive interaction (%)	24.2	8.3	40.5	0.4	23.4	25.0	1054
Praise more than reprimands	79.3	N.A.	N.A.	1.2	76.9	81.7	1054
Actions at the end of the lesson							
Summarised day's lesson	53.4	N.A.	N.A.	2	49.3	57.4	740
Revisited lesson's objectives	22.6	N.A.	N.A.	1.8	19.1	26.1	740
Gave homework	26.6	N.A.	N.A.	1.5	23.7	29.5	740
None of the above	29.3	N.A.	N.A.	1.8	25.7	32.8	740

Source: Quantitative baseline survey (October 2014), TDNA, and classroom observations.

Table 41 Teachers' use of resources in the classroom: Mean estimates

Indicator	Mean	P10	P90	Standard Error	Lower 95% CI	Upper 95% CI	N
Teachers who used (% of teachers observed)							
Textbook	55.3	0	100	1.6	52.1	58.4	1053
Blackboard	95.5	100	100	0.7	94	96.9	1053
Chalk	97.3	100	100	0.5	96.3	98.4	1053
Poster, chart, pictures	8.2	0	0	0.8	6.7	9.7	1053
Improvised materials made by teacher	15.3	0	100	1.2	13	17.7	1053
Resources from local environment	7.1	0	0	0.8	5.6	8.6	1053
Audio	0.1	0	0	0.1	-0.1	0.2	1053
Video	0	0	0	0	0	0	1053
Science equipment	0	0	0	0	0	0	1053
Other equipment	0.6	0	0	0.2	0.2	1	1053
Number of resources used	2.8	2	4	0	2.7	2.8	1053

Source: Quantitative baseline survey (October 2014), classroom observation.

Table 42 School characteristics: Mean estimates

Indicator	Mean	P10	P90	Standard Error	Lower 95% CI	Upper 95% CI	N
Number of teachers employed (Grades 1–6)	12.2	4	26	0.5	11.3	13.2	330
Number of pupils registered (Grades 1–6)	654.8	142	1499	25.2	605.2	704.5	328
PTR	58.7	22.7	106	1.5	55.7	61.7	328
Number of pupils per observed classroom (class size)	41.9	11	82	1.1	39.7	44	1053

Source: Quantitative baseline survey (October 2014), head teacher interviews and teacher interviews. Note: (1) Statistically significant differences between groups are marked with asterisks: *significant at the 10% level; **significant at the 5% level; ***significant at the 1% level.

Table 43 Reported reasons for teacher absenteeism from school

Indicator	Overall	Treatment	Control	N
Reasons for teacher absenteeism (% of head teachers reporting each reason)				
Security/safety concerns	0.6	0.0*	1.2	329
Pay/salary related	33.1	33.4	32.7	329
Low motivation/laziness	23.2	22.0	24.5	329
Social/religious obligations	18.4	15.6	21.4	329
Training	2.3	1.8	2.7	329
Own/family illness	46.1	45.3	46.9	329
Other income-generating activities	7.0	9.6**	4.2	329
Bad infrastructure/conditions	1.8	1.8	1.9	329
Lack of teaching materials	1.5	1.1	1.9	329
Distance/travel time	28.9	32.6*	25.1	329
None	4.6	2.4**	6.9	329
Other reasons	7.1	6.1	8.1	329
Source: Quantitative baseline survey (October 2014); head teacher interviews. Note: (1) Statistically significant differences between groups are marked with asterisks: *significant at the 10% level; **significant at the 5% level; ***significant at the 1% level.				

Table 44 Reported reasons for head teacher absenteeism from school

Indicator	Overall	Treatment	Control	N
Reasons for head teacher absenteeism (% of head teachers who were absent reporting each reason)				
Elections/campaigning	0.0	0.0	0.0	165
Transport	4.7	2.3	7.4	165
Teacher strikes	0.0	0.0	0.0	165
Other mass strikes	0.0	0.0	0.0	165
Own or family illness	55.0	59.3	50.3	165
Late or non-payment of salary	3.1	3.6	2.6	165
Training	17.2	17.9	16.3	165
Meeting or event at LGA/SUBEB	16.5	12.8	20.6	165
Social/religious obligations	8.8	10.3	7.1	165
Epidemic/disease outbreak	0.0	0.0	0.0	165
Bad weather	0.6	1.1	0.0	165
Other reasons	9.9	10.5	9.2	165
Source: Quantitative baseline survey (October 2014), head teacher interviews. Note: (1) Statistically significant differences between groups are marked with asterisks: *significant at the 10% level; **significant at the 5% level; ***significant at the 1% level.				

Table 45 Head teacher – teacher formal meetings topics

Variable label	Overall	Treatment	Control	N
Topics discussed in formal head teacher – teacher meetings (% of head teachers)				
Teacher absenteeism/lateness	69.8	71.2	68.3	326
Pupil attendance	63.1	63.4	62.8	326
Pay/salary	2.8	3.1	2.5	326
Materials	9.3	7.3	11.3	326
School building conditions/repairs	5.3	5.0	5.6	326
Teaching practice/pedagogy	44.9	44.9	44.8	326
Individual students' needs	12.5	12.5	12.5	326
Parents/community	15.4	12.9	17.9	326
Training	2.3	1.9	2.8	326
Professional development	26.6	24.6	28.6	326
Others	8.5	9.1	7.9	326
Source: Quantitative baseline survey (October 2014), head teacher interviews. Note: (1) Statistically significant differences between groups are marked with asterisks: *significant at the 10% level; **significant at the 5% level; ***significant at the 1% level.				

Table 46 Head teacher actions to manage pupil and teacher attendance

Indicator	Overall	Treatment	Control	N
Out of head teachers who took action to improve pupil attendance, action taken was to (%)				
Involve SBMC in finding reasons for non- attendance	75.1	73.3	77.1	326
Discuss with teachers, pupils or parents about reasons for non-attendance	73.1	73.1	73.1	326
Provide financial support	4.2	2.8*	5.8	326
Provide uniforms	6.8	6.0	7.6	326
Provide textbooks, exercise books and stationary	14.3	12.0	16.8	326
Address bullying	0.6	0.6	0.7	326
Address corporal punishment	2.1	2.3	1.8	326
Improve quality of teaching and learning	6.2	5.9	6.5	326
Other reason	11.9	10.9	13.0	326
Out of head teachers who took action to improve teacher attendance, action taken was to (%)				
Rule attendance book at opening time and follow up absences	46.9	44.9	49.1	312
Insist on written absence requests	41.2	43.3	39.0	312

Indicator	Overall	Treatment	Control	N
Complete movement book during school hours	29.4	29.2	29.6	312
Discuss with teachers about attendance	68.5	70.0	66.9	312
Address pay/salary related grievances	9.4	10.4	8.4	312
Address childcare/maternity/paternity related issues	4.5	2.5**	6.6	312
Address issues related to school infrastructure/conditions	1.0	0.0**	2.0	312
Address lack of teaching materials	1.9	1.3	2.6	312
Other reason	10.1	6.5***	13.9	312

Source: Quantitative baseline survey (October 2014), head teacher interviews. Note: (1) Statistically significant differences between groups are marked with asterisks: *significant at the 10% level; **significant at the 5% level; ***significant at the 1% level.

Annex N Supplementary qualitative analysis: Detailed school accounts

The qualitative research design and the rationale for selecting schools in each of the three states – Jigawa, Katsina and Zamfara – for this work have been described earlier in this volume (Section 4). This annex provides a more detailed description of each of the three schools chosen in each state, under the headings: pupils, teachers, SLM, community and parental engagement, and TDP processes.

The aim of this section is to provide a case by case description of the research sites that augments the across cases analysis presented in Volume I.

N.1 Jigawa, high-performing school

Main themes: Weak head teacher school leadership and management, teacher transfer, low teacher motivation, classroom overcrowding

Table 47 Some key quantitative baseline survey data for Jigawa, high-performing school

State	Performance category	English TDNA (%)	Maths TDNA (%)	English literacy pupil test (%)	Numeracy pupil test (%)	# teachers	# pupils	Average class size
Jigawa	High	52	61	18	26	14	980	111

Source: TDP in-service training impact evaluation baseline survey (2014); TDNA and pupil test scores represent marks scored (raw) as a percentage of total marks; # teachers = number of teachers employed (Grades 1–6); # pupils = number of Grade 1–6 pupils registered

This is a medium- to large-sized primary school (1,082 pupils in Grades 1–6, of whom 58% are boys) located on the outskirts of a city. The history of the school is unclear due to a combination of rapid turnover of staff and a lack of school records.

The school has eight classrooms for nine classes, which means the nursery class must sit outside under an area sheltered with metal sheets. Class sizes vary widely due to differing enrolment between grades; however, the overall PTR is high (95:1, including nursery grade).

N.1.1 Pupils

Pupils in this school were generally dressed in plain school uniform and were generally in classrooms during lesson time.

The majority of students reported carrying out household chores such as sweeping, washing clothes and dishes, and fetching water outside school hours. Some pupils attended Arabic school on some evenings, which made them enjoy their Arabic classes at this school more because they felt each act as a booster for the other. One student was also an apprentice at a garage, where he was learning to repair cars in the evening. However, in most cases pupils insisted that none of these activities affected their attendance at school. However, three out of the six girls included in the focus group were absent for at least part of the previous day as a result of needing to return

home at morning break to carry out household chores or care for younger siblings, or due to their own sickness. A large number of pupils were late for the lesson that was observed, increasing the class size from 84 at the beginning to 126 by the end.

Pupils described unhappy children as being physically hit by their teachers and parents for late attendance at school. By contrast, happy children were helped by the parents with their homework.

N.1.2 Teachers

There are nine female teachers and 4 male teachers in this school. Two of the three TDP teachers were relatively young, enthusiastic men and NCE-qualified. Most teachers in the school had also received ESSPIN training, which they felt was similar to TDP.

Teachers stressed the importance of preparing lessons before going into the classroom, and used the ESSPIN lesson plan to guide their lessons where possible. The ESSPIN lesson plan was so long that the head teacher changed the school lesson timetable to allow for double periods during which the ESSPIN lesson plan could be taught. The head teacher reads and approves lesson plans prepared by the case study teacher. The head teacher noted that 'successful' teachers teach step by step and ensure that their pupils understand each step before proceeding to the next. He explained that sometimes the teacher will teach one topic for a whole week in order to achieve this. The head teacher also believed that the use of teaching aids was necessary to teach effectively. However, as noted above, teachers complained that there was a lack of teaching aids that enable them to teach the curriculum effectively.

Despite these claims to be adopting more child-centred teaching techniques, students said the main activity they undertook in their most recent class was copying notes written by the teacher on the blackboard. However, during the lesson observation the teacher did move around the class to check on the quality of the notes being taken. The teacher also grouped pupils to work together and used flashcards with the names of animals to facilitate learning. However, the large class sizes meant that the groups were too large for all students to participate. High-performing groups were asked to read their work aloud to the class.

Teachers complained that classroom overcrowding made control of pupils difficult and group activities became unruly, leading one teacher to ask 'how can you teach a class of over 200 pupils?'. They also noted that parents take their children out of the school to farm. These two issues demotivate teachers, who turn to each other for support and advice.

Other teachers were demotivated by the lack of a staff room for teachers, pupils attending school without stationery, and delayed promotions whereby younger teachers 'leap-frog' older teachers in terms of responsibility and pay.

Despite the alleged suitability of teaching careers for women, the LGEA still received many complaints from head teachers regarding the late attendance of female teachers. In this school, on each day a group of three to five female teachers often sat together outside whilst they tended to their young children. They left several classes unattended in the process, sometimes for more than half a school day. The head teacher did not appear concerned by this behaviour, and did not refer to it as being a challenge in conversations.

Lessons are taught in both Hausa and English, and during the lesson observation students were far more engaged when questions were asked in Hausa. Tests are administered in English; however, pupils have difficulty understanding English written text.

N.1.3 SLM

The current head teacher was appointed to the position of head teacher in April 2015, having previously been deputy head teacher in the same school. The school has had three different head teachers over the course of one year: the previous two were both promoted to positions in the LGEA. The head teacher feels that he is respected and valued by his teachers, pupils and members of the local community. He explained that parents' decisions to transfer children from other schools to this school were evidence of this value, and that parents visit to thank him in person. All teachers felt valued by the head teacher, and they felt that he provides good leadership of the school.

The head teacher feels he has no authority to discipline teachers. He does not report teacher absenteeism to the LGEA, and instead leaves inspectors to explore the issue and to take it further when they visit the school. Despite the obvious challenges of unattended classrooms, he claimed that he does not 'report teachers because I don't have any problems with them'. TDP teachers did not feel they were rewarded for positive performance, beyond receiving verbal praise. Teachers said that they were motivated by lesson observations by the head teacher, since it makes them feel that they are valued and also helps them to improve their teaching technique.

Despite his insistence that he had no problems with his teachers, the extent of teacher classroom absenteeism in this school was clear evidence that he was either unwilling or unable to influence teacher behaviour. Since the largest cause of teacher classroom absenteeism was maternity commitments to care for young children, it is possible that such absenteeism was seen as permissible (or at least that it was unfair to criticise mothers for not fulfilling their teaching commitments).

The head teacher is not aware of the recruitment process, and complains that some teachers were transferred away but not replaced. He had previously requested an Arabic teacher to replace one of these, which the LGEA subsequently provided. He has requested additional support, though these requests have not yet been granted and he claimed that '[he doesn't] even follow up or call to ask … because I don't have the right.'

However, the LGEA finds it difficult to meet demands for additional staff from school, since it claims it is no longer responsible for the recruitment of teachers, only for allocating them within schools in the LGA. The SUBEB was reported to not be employing new staff quickly enough, and retired staff have gone un-replaced. Furthermore, the LGEA's recruitment recommendations to the SUBEB are rarely accepted, with only five out of 40 recommendations actioned by the SUBEB. This treatment, together with the fact that the LGEA does not have a permanent office of its own, combine to make the LGEA feel degraded.

Given the lack of control over recruitment and pay, the LGEA officials noted that the main disciplinary mechanism available to them was the transfer of teachers from urban to remote rural areas within the LGA.

N.1.4 Community and parental engagement

Opinion was divided on the issue of community relations. Most teachers felt that they were valued by the local community, as indicated by the fact that they were invited to local weddings and events. However, the case study teacher admitted that he did not feel the community values teachers, since there is a perception that teaching is a 'fall back' career for those without any other job options. He did clarify that teachers generally have a good relationship with pupils' parents.

The head teacher believed that it is important to discuss challenges and problems with the SBMC, and had been reading ESSPIN guidance on how to run the SBMC. Having only been appointed three months previously, the new head teacher had yet to meet with the SBMC, but was compiling a list of existing parent–teacher association (PTA) and SBMC members to arrange a meeting. He noted that 'a good relationship with the community will encourage pupils' enrolment into the school'. Traditional leaders from the local community also visit the school occasionally. The previous head teacher was very sociable, and had encouraged parents to send their children to school via the SBMC, and had organised meetings with the teachers for them to share ideas and advice.

N.1.5 TDP processes

The TDP teachers receive all the intended interventions. The head teacher praised TDP as very good and said it was improving teacher training techniques. He suggested that more teachers should be involved.

The teachers said that the amplifier helped them to project their voices in congested classrooms. They were also using the amplifier in innovative ways: for example, one teacher recorded stories on his phone and then connected the phone to the amplifier to play them to the class. This apparently helps pupils to improve their listening skills. However, when this was demonstrated during the lesson observation, most students appeared not to be listening.

They believed teachers generally articulated themselves freely during cluster meetings, but that some were afraid of making mistakes. They felt motivated to teach more effectively after going to cluster meetings. The teachers suggested that cluster meetings should be held more often (every three weeks) and should cover other subjects as well.

The TFs noted that there was a shortage of materials for training, such as the TDP video clip, flip charts and A4 paper. The case study teacher reported that he did not receive any materials from the most recent cluster meeting. Furthermore, there were complaints that the tablets, trainer in the pocket, and typed handouts did not share a similar structure and often made references to materials (e.g. textbooks) that teachers did not have access to.

However, TDP teachers were aware that non-TDP teachers felt excluded and did not want to learn new teaching techniques from the TDP teachers, despite attempts by the TDP teachers to share resources.

Both of the TFs were employees of the LGEA and worked with ESSPIN, which apparently helped in their recruitment to work on TDP. However, the LGEA official reported that the LGEA staff – particularly QAOs – should be more involved in the implementation of TDP, since often the main

government inspectors have no knowledge of the teaching techniques TDP is trying to promote or the stage of TDP implementation.

TFs complained that they found it difficult to visit all the six schools they were each responsible for because some are located in very rural areas with poor accessibility. They requested that vehicles be provided to make these journeys easier. They also requested that they be given feedback following their report to TDP on their school support visits.

The TFs also suggested that all TDP teachers should receive both literacy and numeracy training, so that any of the TDP teachers can fill in for absent teachers or those that are transferred away. Teachers felt that TDP should provide school infrastructure to facilitate adoption of the next teaching techniques.

Finally, there had allegedly been complaints that the teachers did not receive their allowances for attending cluster meetings immediately, unlike in the case of ESSPIN.

N.2 Jigawa, typical school

Table 48 Some key quantitative baseline survey data for Jigawa, typical school

State	Performance category	English TDNA (%)	Maths TDNA (%)	English literacy pupil test (%)	Numeracy pupil test (%)	# teachers	# pupils	Average class size
Jigawa	Typical	28	37	13	15	8	350	36

Source: TDP in-service training impact evaluation baseline survey (2014); TDNA and pupil test scores represent marks scored (raw) as a percentage of total marks; # teachers = number of teachers employed (Grade 1–6); # pupils = number of Grade 1–6 pupils registered

During the qualitative fieldwork we found seven teachers and four classrooms in the school. Three other classes (including nursery) were being taught outdoors.

N.2.1 Pupils

Boys substantially outnumbered girls in the school. According to the head teacher this was the result of the community not valuing girls' education, although he also claimed this situation had improved recently. However, a previous head teacher had been found to have sexually abused girls in the school. The absence of any female teachers in the school could also be a factor in the low enrolment of girls.

Around 200–250 children were in attendance during the qualitative fieldwork, compared with the total enrolment of 350.

In FGDs, most boys said they sometimes missed school because of going to work with their fathers on the farm. In particular, they tended to miss school during the rainy season. Children also said they worked at home on household chores, such as fetching water, but that this happened after school hours, whereas farm work happened during school hours.

Boys said they liked their teachers; there was formerly one teacher who beat them but he was transferred to another school.

Children explained that teachers sometimes did not come to the class, in which case they would simply wait, or they could read their notes from the previous lesson. The children could demonstrate knowledge of several English words in speech but could not read any. They could not identify the words in their notebooks that related to the subject of a lesson, or distinguish words from numbers. When we started to read from their notes, they would remember what the lesson was about and respond by telling us more about it, indicating that they were understanding at least some of the lesson and partially memorising the content, but struggled particularly with written notes in English.

Children in the focus groups wished that they could improve the ceiling, walls, chairs, and blackboards, and increase the number of classrooms.

The head teacher described a case of a 'brilliant' boy pupil who was removed by his parents to study in the almajiri school following the divorce of the parents. He suggested that step-parents do not always give adequate support to their children.

The children in our lesson observation appeared eager to learn, copying in their notebooks and competing to answer questions. Not all had uniforms and the school did not appear to be strict about this.

N.2.2 Teachers

Teachers said they were demotivated by infrastructure problems in the school, including poor blackboards and dirt floors. The lower classes were taught outdoors, under a tree, and during the rainy season or *harmattan* they would either not be taught, or would be distributed to the other classes. Most of the teachers came very late to the school – around 9am – and the reasons for this were not fully clear.

The head teacher argued that teachers in this school use appropriate teaching methods and improvise a lot, including using the 'play way', drama, and placing children in groups. During our lesson observation we observed some diversity of methods, including pupils being brought to the front to demonstrate how to solve mathematics problems, the use of flashcards, counting in groups and counting all together. However, we also observed a lot of emphasis on writing notes on the board that the children could then copy. Teaching was largely in Hausa, even for higher grades. Teachers said they understood the curriculum very well, but that there were problems with the level of the curriculum being difficult for the children.

N.2.3 SLM

The situation where the previous head teacher had been found to have sexually abused girls in the school had, according to the current head teacher, been resolved without involving the police. Community members had written to the school, and then the LGEA sent a committee, which sat with the father of one girl who had alleged abuse, and other members of the community. Together they decided that the head teacher had indeed committed the abuse, demoted him to the status of regular teacher, and sent him to another school. The father then agreed to drop the matter. Another case, of a teacher who had been coming to school drunk, was also handled with a transfer to another school.

Like the teachers, the head teacher said he was demotivated by the inadequate infrastructure in the school.

The head teacher felt that teachers in the school respected him. He claimed to carry out lesson observations 'from time to time' but had not done so this term.

School inspection was felt by teachers to be punitive rather than supportive in nature. Our case study teacher cited one particular case where inspectors had come early in the morning and found teachers absent, resulting in the attempted transfer of the teachers to another school (a move which was stopped after community members intervened).

There were large numbers of textbooks in the school, but these appeared to be used only by the teacher and not handed out to the pupils.

Teachers and the head teacher said they were largely satisfied about issues of salary and promotion. The teachers we interviewed were happy being teachers and did not aspire to any other profession. However, our case study teacher also worked in an almajiri school, and worked as a carpenter – he argued that this extra work outside school hours allowed him to live more comfortably when his salary as a teacher was not paid on time.

N.2.4 Community and parental engagement

The school's head teacher was from the local community, while other teachers in the school came from slightly more distant villages. The head teacher felt that the local community respected him. He gave as an example the aftermath of the sex abuse case, where he had persuaded the fathers of the abused girls to continue sending their girls to school after the former head teacher had been removed from the school.

The current head teacher, and other teachers, had also been posted to a different school in the past but (according to the head teacher's own account) had returned after the SBMC and other members of the community protested and met with the LGEA officials, asking for him to be brought back to the school. Despite this evidence of SBMC activity, the SBMC did not appear to be active in the regular functioning of the school, and had not met recently. However, members did apparently visit the school and were aware of its poor state of repair, but did not know what they could do to address this. The SBMC had also been active in persuading a father whose son had dropped out to send the boy to school again. It made occasional donations to the school, such as empty cement bags for use as makeshift prayer mats, but appeared unable or unwilling to make a serious impact on the school's poor infrastructure or lack of learning materials.

Community members said they appreciated the teachers of the school and could see evidence that their children were learning. They said they could report teachers who perpetually arrived late to the head teacher.

N.2.5 TDP processes

The school had not in fact been included in TDP, despite being on the original list of intervention schools and included in the quantitative baseline survey. The head teacher had not contacted anyone to enquire about the reasons for this. Local government officers clarified that the school

was not on their list of TDP schools, but we were unable to determine the reasons why it had been included on our list and yet not in the intervention.

ESSPIN was also active in this LGEA, and the TDP TFs were also ESSPIN officers, and split their time between the two programmes. They told us this was not a problem as activities for the two programmes happened at different times. The facilitators had been selected, according to the facilitators themselves, on the basis of recommendation from ESSPIN.

N.2.6 Summary

In some ways the school could be said to be functioning relatively well given its remote rural location, and social and economic context. For example, there are reportedly very few female teachers in the area, and a widespread problem with girls not attending school, so even the small number of girls who attended the school may represent a small improvement on other schools in this context. Learning levels of teachers and children were also somewhat higher than in other comparable schools. However, the school had severe infrastructural problems, issues with teacher lateness and non-attendance in class, and erratic attendance by pupils. Local and state-level government were reportedly paying little attention to its needs. It is not clear whether it would be within the means of the local community to improve the school's infrastructure, for example by lobbying local government or providing contributions in kind. The exclusion of the school from TDP intervention suggests some problems with either the school selection process or the process of managing information about which schools have been selected.

N.3 Jigawa, low-performing school

Main themes: Low teacher incentives and motivation, high pupil and teacher absenteeism, classroom overcrowding and poor infrastructure, weak SLM, high levels of household poverty pupils

Table 49 Some key quantitative baseline survey data for Jigawa, low-performing school

State	Performance category	English TDNA (%)	Maths TDNA (%)	English literacy pupil test (%)	Numeracy pupil test (%)	# teachers	# pupils	Average class size
Jigawa	Low	2	29	16	24	5	409	20

Source: TDP in-service training impact evaluation baseline survey (2014); TDNA and pupil test scores represent marks scored (raw) as a percentage of total marks; # teachers = number of teachers employed (Grades 1–6); # pupils = number of Grade 1–6 pupils registered

This is a medium-sized primary school with an enrolment of about 409 pupils and attendance, on the day of the qualitative survey, of about 343 from pre-primary to Grade 6. The school is located along an arterial, paved, road within an hour's drive of Dutse, the capital of Jigawa state. The school has six classrooms for six classes, which means the pre-primary class has to sit outside under a tree, and the teachers also have to sit outside on a bench in their free periods. Class sizes vary widely due to differing enrolment between grades; however, the overall PTR is high (82:1, including nursery grade).

The head teacher particularly commented with regard to the number of classrooms that the school was in dire need of additional rooms, especially for the pre-primary pupils who are around 275 in a class, all sitting under the tree out in the open. In addition, when it rains some of the classrooms are unusable due to leaky roofs – as a result, sometimes all pupils have to be accumulated into a single classroom. Not all classrooms have furniture or useable blackboards, some classrooms have bats on the ceilings and floors are rough and damaged, making them difficult for pupils to sit on in the absence of furniture. There is also a reported problem of some 'local thugs' breaking into the head teachers' office at night and stealing a lot of things. The head teacher reported a severe shortage of learning materials and, on the day of the qualitative survey, the school had apparently run out of chalk altogether.

N.3.1 Pupils

Teachers and the head teacher complained about high pupil absenteeism in this school, especially during the rainy season and harvest season when children are sent to work on the farms. To alleviate this, the head teacher had reportedly spoken to the community members and parents, asking them to send their children to school. According to the head teacher this does not seem to have been terribly effective. Parents allegedly often came to school in the middle of the classes and withdrew their children for that day to send them to the farm.

In researchers' interactions with pupils the pupils described their aspirations for their future – some said they wanted to be doctors, some veterinarians, some pilots, some fishery experts and some even school teachers. When asked what they would do if they were granted magical powers, they almost unanimously said they would fix the school infrastructure by adding more classrooms, fixing the ceiling and 'changing the structures'.

N.3.2 Teachers

There are five teachers in this school, including the head teacher, and four of them are in receipt of TDP training. One of the TDP teachers was absent for all three days during which the researchers were in the school and the head teacher was unsure about the reason for her absence or when she would be back. She did attend the TDP cluster meetings during this period, but not the school.

This school was chosen from among the lowest performing sampled schools in Jigawa, in terms of teachers' TDNA scores. Ironically, the head teacher and teachers claimed they knew 'everything in the curriculum'. The head teacher did mention, however, that the teachers were not very hard working and that they 'don't apply themselves'. At the same time he stated that his teachers were getting proficient in the 'play way' of teaching, and were gradually using more group work.

The case study teacher discussed, among other things, his various income-generating activities, which he pursues outside the school but which he claimed did not interfere with the time he needs to devote to his job as a teacher (even though this is not easy). These included animal husbandry, metal works, and farming. According to this teacher it is parents who are mainly responsible for the poor performance of the pupils in the school because they do not contribute time to teach their children at home. The teacher laughed when researchers asked if he would ever send his own children to the school – perhaps this meant he would not. They presently went to a private school closer to the nearby town. He was happy with the progress he had achieved

that week – he had taught his pupils well and felt they had understood everything well, and he had also been observed by the head teacher that week.

N.3.3 SLM

One of the main issues the head teacher discussed was the way teachers were recruited into his school. According to him, 'lazy' teachers who had personal connections seemed to make their way to his school and he had no say in this process. He found it very difficult to secure and retain talented teachers in his school and felt quite powerless in this regard.

He also did not have much to incentivise teachers to do well, in his view. One of the limited ways to influence teacher performance was through lesson observations and giving feedback to teachers – constructively discussing their strengths and weaknesses, both with the teachers themselves and with the LGEA officials. He believed that the only way to exercise authority and manage teachers is through words of appreciation, or to advise them not to be absent, and to use appropriate teaching methods in the class.

Otherwise, the head teacher struggled to find ways to take action against his teachers for any wrongdoing or poor performance – 'because if you cannot stop their salary and you cannot terminate their appointment or suspend them what else will you do to discipline them?'

The head teacher talked about a special cadre of school inspectors constituted by the former governor, Sule Lamido, called Social Mobilisation Officers, whose job is to visit schools and report their findings on the working of the school straight to the governor – as such, they have the power to fire a teacher or head teacher if they are 'not comfortable with your style'. However, these officers seem to provide no constructive feedback to teachers or to the head teacher.

N.3.4 Community and parental engagement

According to the head teacher, the school's SBMC had been very helpful in a number of ways and was doing its best to deal with some teething issues, like pupil absenteeism and school infrastructure.

The SBMC had taken care of the school by repairing the school's borehole from time to time, since it was the only source of water in the school. Sometimes it also procured writing materials for the pupils, and brooms to clean the school. The SBMC had also constituted a team of local village volunteer youth to act as stewards, to help pupils cross the road on their back to and back from school. For other materials, like ablution pots, and sports items such as footballs, etc. the school usually asked parents to contribute around five Naira towards the school fund.

The parents of children who attended this school were described by the head teacher as being very poor and unable to afford the most basic writing materials or uniforms for their children. Some of them were petty-traders, some hawkers, some animal rearers, some butchers and some fishermen. High levels of household poverty meant children often came to school hungry or with little or no pocket money for food. The head teacher asked rhetorically: 'what can you expect from them?' in terms of learning, since they come to school hungry. A number of times the head teacher or other members of school staff would assist some of the poorest pupils with food or a uniform.

When parents came to the school to discuss how little their children seemed to be learning in the school, the head teacher and teachers claimed it was the parents' fault for not sending their children to school regularly, for allowing them to be habitual latecomers and for enrolling them in almajiri schools, which sometimes led to them missing regular school. It was therefore not surprising to the school staff that the children of these parents were unable to learn much in school.

N.3.5 TDP processes

The head teacher mentioned that he was unaware of the reason why his school had been selected as a TDP school, though he was pleased with the outcome. He guessed it was perhaps because the LGEA inspectors deemed his school to be a high-performing one (mainly in terms of 'keeping proper records and punctuality') and worthy of TDP's intervention. The head teacher also wondered if his school had been picked because it was also an ESSPIN school, since 'ESSPIN is just like TDP'.

From this school three male teachers, including the head teacher, and a female teacher were selected to participate in TDP. The female teacher was absent from school for all three days of the survey team's stay in the school; she did attend the cluster meeting for two of these days but did not attend any lessons in school during this period. She lived in the nearby town and not in the community adjoining the school.

The head teacher suggested that in a small school like his, with just five teachers, having two teachers missing from school for two consecutive days a month for cluster meetings was detrimental to managing the school. Instead of running the cluster meetings from 9am–4pm he suggested splitting it up into three days and spending about four hours each day, and attendees being allowed to return to school for the rest of the day.

It was alleged that some state-level school inspectors had made rounds of the school and had discouraged the teachers from using the TDP materials, especially the audio-visual materials, since the inspectors were unfamiliar with these pedagogical methods. The LGEA officials, on the other hand, who had some awareness of TDP and its training model, had encouraged teachers to use them.

N.4 Katsina, high-performing school

Main themes: Hostility between TDP/non-TDP teachers, teacher selection issues, relatively strong SLM, relatively good infrastructure, patronage from well-connected community members, parental aspirations, continuous pupil assessment

Table 50 Some key quantitative baseline survey data for Katsina, high-performing school

State	Performance category	English TDNA (%)	Maths TDNA (%)	English literacy pupil test (%)	Numeracy pupil test (%)	# teachers	# pupils	Average class size
Katsina	High	39	88	14	30	17	978	39

Source: TDP in-service training impact evaluation baseline survey (2014); TDNA and pupil test scores represent marks scored (raw) as a percentage of total marks; # teachers = number of teachers employed (Grade 1–6); # pupils = number of Grade 1–6 pupils registered

This is a medium-large primary school in Katsina, roughly an hour's drive from Katsina city. It has about 1,000 pupils enrolled, from nursery to Grade–6, with around 20 teachers on the register, including the head teacher. As such, it has a PTR of 58:1 and an average class size of 40 pupils per class. There are 14 classrooms for 14 grades and arms, one library and one IT centre, both of which appeared fairly well-stocked but usually remained under lock and key, and were not frequently accessible to pupils. The look of the school is that of a well-endowed school with good infrastructure, which indeed it is. The classrooms have clean and floors are tiled, all classrooms have whiteboards, and each pupil has a chair and desk for himself/herself. As such, classes are not overcrowded.

N.4.1 Pupils

Pupils in this school are required to sit for formative and continuous assessment every Friday, which involves a short class test on the subjects being taught. The head teacher claimed scores from these assessments were used for commending well-performing pupils and identifying the weak ones and discussing their progress with their parents.

This school claimed to have among the best results in the primary school leaving exams in Katsina State, and was seen as a 'model' or trophy school. It was mentioned several times by the Education Secretary as one of the best schools in the state. Pupils from this school regularly achieved top positions in local debate and quiz competitions at the LGEA level.

All pupils in this school were in uniform, since the school, and especially the head teacher, was very particular about pupils wearing clean uniforms every day to school. To instil this discipline in pupils the head teacher often picked on pupils in the morning assembly who were in unkempt uniforms (or were not in uniforms at all) and punished them in front of the entire school, to make an example out of them and so deter others from doing the same, even though it was probably extremely demotivating and embarrassing for the pupil being punished. Other types of corporal punishment were also regularly practised in this school – including forcing children to climb the school's flagpole and latecomers being punished in front of the entire school, etc.

For pupils to be enrolled in the school, the head teachers had instituted a rule that only pupils brought to the school by their parents or grandparents (no older siblings, neighbours, etc.) and with a valid birth certificate would be considered for admission.

As a cumulative effect of all these measures (strict rules around uniform, the need for pupils to have a birth certificate and accompanying parents to admit them to school, etc.) it appeared that

the lives of relatively poor pupils in this school were not easy and were possibly marked by discouragement, punishment, and academic and social exclusion, thereby leading to an almost natural selection of pupils from certain types of backgrounds who are admitted to, and prosper in, this school.

N.4.2 Teachers

All teachers in this school are NCE holders, barring one who is presently on long-leave, to complete his NCE.

Even though none of the TDP teachers were from the local community they claimed to have the respect of the community members: 'Yes people respect us a lot in the community we work in, even though we are not from this community but they always like to involve us in their community, especially during special occasions like weddings and naming ceremonies.'

The two TDP teachers who were interviewed mentioned that they were initially planning to pursue a career in the Nigerian armed forces but eventually decided to take up a career in teaching because of the way a teacher is respected in society and is seen as a dignified figure.

The teachers in this school listed their key constraints as follows. They complained about the negative and demotivating influence of the local supervisors/inspectors, who invariably end up giving them 'negative commendation' after lesson observations, despite how hard the teachers try to do their best in class. The teachers also mentioned the difficulty of accessing the school on their motorbikes because the road leading to the school is unpaved and is in a bad condition. There was also the issue of teachers' accommodation – since none of the teachers are indigenes of the local community, they explained that it would really help if they were provided with accommodation in the local community.

N.4.3 SLM

The head teacher commented that while they were comfortable with their existing infrastructure, over the next three to five years he expects there will be need for additional blocks of classrooms due to the rapidly increasing enrolment in the school. Regarding textbooks, the head teacher mentioned that there were frequent shortages of basic science and Islamic education books, as well as materials for teaching arts and crafts to the pupils – some of these he purchased personally to supply to the school. The head teacher still maintained fairly active teaching engagements in the school.

The head teacher seemed to have well-established connections with officials at the LGEA, due to which a number of the issues raised by him seemed to have been addressed: 'I am just very lucky whatever I reported they will follow it up and take action against it.' For instance, the head teacher had complained about insubordination by a teacher in the school and about another teacher who demonstrated curtailed instructional time in the class – in both cases the head teacher claimed the LGEA had taken appropriate 'action against him' – namely transferring him to another school.

The head teacher expressed the common complaint of feeling powerless when it came to teacher recruitment in this school.

The head teacher mentioned that, according to him, one of his most important contributions to the school has been bringing discipline to the school: ensuring pupils turn up in a (clean) uniform, that they arrive on time in the morning, and that they return to their desks promptly after lunch break – and generally that they maintain decorum in the classes and school compound during school hours. He practised a range of physical punishments on pupils who erred on some of these fronts – e.g. on one of the survey days, pupils were seen climbing the school's flagpole as a punishment.

N.4.4 Community and parental engagement

Members of the local community, the main catchment area for this school, included senior figures in the Katsina SUBEB and UBEC staff at the federal level – and some of these well-connected individuals were patrons of this school. The head teacher and teachers claimed that the excellent learning resources and play materials, furniture and other infrastructure could directly be attributed to this school's connections to these powerful individuals in the state and federal education administration. This included receiving 'direct intervention' grants from the UBEC.

The head teacher claimed that this school has a fairly effective SBMC which, along with the PTA, was actively involved with the school administration in tackling issues like pupil absenteeism and late arrival. The head teacher and teacher also claimed that the school staff had a very cordial relationship with the SBMC and PTA: 'Last year during Ramadan period the PTA brought sugar, rice and other things just to appreciate us.'

Most of the parents of this school's pupils were employed in the LGA offices. The head teacher claimed that, over time, members of the community, including the parents, had seen pupils graduate from the school and get jobs either in the local government or elsewhere, and hence this had reinforced their aspirations for their own children who are pupils in the school currently, and thus it prompted them to invest in their education, both materially and non-materially. It appeared that parents and community members had, over time, seen the schooling-employment nexus fructifying, which had strengthened their expectations in regard to the school and to their children's education.

N.4.5 TDP processes

As per standard, four teachers, including the head teacher, had been selected in this school for participation in TDP. The head teacher was not sure why his school had been selected and wondered if it was because his school is seen as a model school in the LGEA.

One of the teachers of the school, the case study teacher for this school, complained that he was first selected by the Education Secretary and had even participated in the quantitative baseline survey and appeared for the TDNA, but was then replaced by another teacher in the school by the head teacher. This replacement, the replaced teacher claimed, was unfair and had left him disgruntled and demotivated, and he was keen for the survey team to report his grievance against the selection process to TDP. One of the teachers selected into the programme from this school was away on long-leave to secure his NCE; however, no replacement was selected to receive the TDP training in lieu of him.

This school also displayed implicit, and apparently even explicit, friction between TDP and non-TDP teachers – largely driven by the frustrations of the non-TDP teachers about not been selected into the programme. The non-TDP teachers claimed no learning materials from TDP training were being shared by the participant-teachers with their colleagues. The non-TDP teachers expressed a keen interest in receiving TDP training as well. After the departure of the survey team, there were reports of tense arguments between the TDP and non-TDP teachers in the head teacher's office.

N.5 Katsina, typical school

Main themes: Pupil absenteeism, parental engagement, strong head teacher SLM, small class sizes

Table 51 Some key quantitative baseline survey data for Katsina, typical school

State	Performance category	English TDNA (%)	Maths TDNA (%)	English literacy pupil test (%)	Numeracy pupil test (%)	# teachers	# pupils	Average class size
Katsina	Typical	20	47	33	47	19	1600	33

Source: TDP in-service training impact evaluation baseline survey (2014); TDNA and pupil test scores represent marks scored (raw) as a percentage of total marks; # teachers = number of teachers employed (Grades 1–6); # pupils = number of Grade –-6 pupils registered

This school is a medium-sized primary school (648 pupils in Grades 1–6, of whom 53% are boys), located on the outskirts of a large town. The school was founded in 1985, and has grown rapidly since 2000, with additional classroom blocks constructed in 2006, 2010, 2011 and 2013, with support from the (federal) UBE, state government, LGA, and Millennium Development Goals (MDG) initiative. Partly as a result of this rapid expansion, the school has relatively small class sizes (approximately 30–50 pupils) and a low PTR (25:1).

There were 18 classrooms in the school, though not all were used for teaching: one was damaged in a fire, another was used as a storage room, and a third was converted into a library. The quality of classroom infrastructure varied significantly. Some classes lacked a blackboard (instead, teachers wrote on sections of the wall that were painted black), whereas classrooms for the older classes were recently equipped with whiteboards and marker pens. The school's library contained a large number of textbooks, though most had clearly not been used in a significant period of time and classes were often taught without textbooks.

N.5.1 Pupils

Teachers in the school have become demotivated by the issue of pupil absenteeism, which was frequently attributed to a lack of adequate support from parents. Pupils often come to school late, and frequently do not return from the morning break, due to the requirement to earn money for the family through hawking. The head teacher has met with the Education Secretary and some parents to discuss solutions to this problem. One compromise has involved allowing pupils to sell goods (e.g. mangoes) inside school grounds during the mid-morning break. This means that more pupils remain on school grounds and has allegedly led to parents being more willing to send their children to school, thereby increasing enrolment.

N.5.2 Teachers

The school has 33 teachers, of whom a significant majority are women and are NCE-qualified. The lesson observed as part of the case study used textbooks (roughly one between two pupils).

Classes were generally better attended by teachers than in other schools included in the study; however, unattended classes were still common. A large part of the lesson involved the teacher reading directly from the textbook, or using it as a prompt. The consequent lack of eye contact and movement likely led to a less engaging teaching style that limited the teacher's efforts to build a rapport with pupils. The teacher distributed white A3 paper to students for them to use in groups in a writing exercise, practising the present perfect tense. However, some groups were confused and required close supervision from the teacher to understand the task they had been assigned. The teacher walked around the classroom to support students whilst the activity was underway. Some students seemed to be mostly copying material from the textbooks rather than coming up with their own examples, as was intended. The quality of pupils' work varied significantly: a significant proportion of the written work was largely illegible.

In contrast to the lesson observation, the FGD discussions with pupils revealed that the main activity during the class they had that day was copying text written on the blackboard. However, pupils complained that the blackboard was cleaned immediately at the end of the lesson when the teacher left the class, meaning that they were unable to finish copying the text into their workbooks. They also complained that they had already been taught that day's English lesson, suggesting that there is some degree of repetition of lesson plans.

Both TDP and non-TDP teachers were perhaps unusually willing to admit that there was room for improvement in their teaching technique: both groups ranked 'poor teaching technique' as a moderate-to-high constraint on effective teaching during the group interviews. Non-TDP teachers were self-conscious of not being trained in 'new' teaching methods used in classes taken by TDP teachers. This seems to have been worsened by the fact that non-beneficiary teachers felt that TDP materials were not shared with them.

Teachers claimed that they were motivated to become teachers because they liked the idea of improving the children's future. However, some teachers in this school were frustrated at delays in promotion and the political influence of 'godfathers' in the promotion process. Teachers often had aspirations of leaving to pursue further studies at degree (usually a B.Ed.) or master's level.

Teachers were demotivated by the frequency of teacher transfers between schools. One reported being transferred to 12 schools in the course of their career. Teachers felt that the system of transfers is used by the LGEA as a means of punishing them. This caused resentment, since 'the teacher's offence will not be well investigated before they are transferred'. The LGEA official denied the claim that transfers were associated with punishment. According to this official teachers are ideally transferred every two to three years according to subject-specific demands from schools, and to prevent a teacher from becoming 'too familiar with his environment'.

Teachers use both English and Hausa in their lessons, with difficult words translated into Hausa for pupils to understand. However, pupils in the school are allegedly assessed in English.

N.5.3 SLM

The current head teacher was appointed as a teacher in the school in 1999, before being transferred away and then returning as assistant head teacher, then becoming finally head teacher in 2012. He appeared generally well-motivated compared to other head teachers and is pro-active in finding solutions to the school's problems. For example, he appeals to his local representative in the state assembly to mobilise support and resources to improve school infrastructure and resources, and maintains a generally good relationship with the LGEA.

The head teacher noted that teachers seeking work experience after completing their NCE have previously taught in this school on a voluntary basis. He wished he was able to recruit these teachers, since they are generally enthusiastic and effective. Although he lacks a formal opportunity to influence recruitment, the head teacher has written to the LGEA to request that teachers with qualifications in subjects that the school is lacking are provided.

TDP teachers felt valued and respected by the head teacher. They ranked 'lack of leadership from head teacher' as the second-least binding constraint faced by teachers in the school. Non-TDP teachers also reported being valued by the head teacher; however, they ranked the 'lack of leadership' as the fifth-highest binding constraint.

The head teacher identifies 'successful' teachers through the exam results of the pupils they teach, through ongoing assessment of pupils based on the scheme of work, and lesson observations. The head teacher carries out lesson observations once or twice a day and follows up with feedback afterwards in private, ensuring that he combines criticism with praise. Teachers agreed that such observations encourage teachers to work harder. Teachers are praised by the head teacher for achievements during staff meetings. Teachers are also commended when they take pupils on excursions organised by the LGEA.

The head teacher disciplines teachers when they are late by not allowing them to sign the attendance logbook. This is a way of indirectly reporting late teachers to inspectors, who check the logbook when they visit. For offences not related to teacher lateness, he asks teachers to write a description of the offence and sign in a separate 'black book' for such offences. The head teachers threatens teachers that he will show the book to inspectors. He believes that these two techniques are effective ways of incentivising the positive performance of teachers, and are a substitute for more formal mechanisms by which the head teacher can discipline teachers. One of the case study teacher's photos showed the head teacher reviewing teacher records to ensure they were correct.

The head teacher was demotivated by pupil absenteeism and by teachers arriving late to school. Teachers did not admit to late or non-attendance at schools, and ranked teacher absenteeism as amongst the lowest constraints on their effectiveness.

N.5.4 Community and parental engagement

TDP teachers did not feel they were valued by the parents of many students. They were demoralised by the fact that the parents do not send their children to school on time or encourage them to return after the morning breakfast break. Teachers ranked the consequent high pupil absenteeism as the third-highest constraint they faced in being effective teachers. Parents also fail to provide their children with writing materials or breakfast, which further demoralises TDP teachers. As a result, both TDP and non-TDP teachers ranked the lack of support from parents as

the second-highest constraint they faced. They were also frustrated with parents that challenged them when the teachers disciplined their children. This matches non-TDP teachers' views that the treatment of pupils was one factor in influencing their relationship with the community.

Teachers often report poor performance of pupils to the head teacher, in particular when pupils had inadequate books and stationery to allow them to benefit from school. In such cases the head teacher then talks to the parents of these children to help resolve such issues.

The SBMC meets at the end of every school term, and is constituted by the head teacher, local community leaders, and (generally richer) parents. The SBMC contributes to the school by getting community members to contribute to repairs. One example of this is the provision of welding services to repair broken chairs. According to one senior LGEA official, schools in semi-urban areas receive more support from SBMCs than those in rural areas, due to having more educated and engaged parents.

N.5.5 TDP processes

The head teacher claimed that all schools in the area receive TDP support, and that they were selected on the basis of good performance. He seemed aware of the official TDP criteria for selecting beneficiary teachers.

The TDP teachers said that the only materials they had received were the 'trainer in the pocket' phone, which they use almost every day. One of the case study's photos showed a non-TDP teacher adopting TDP teaching techniques by grouping pupils' desks and encouraging their participation in the class. The case study teacher allegedly shared the training ideas and tips they had picked up in cluster meetings with other teachers during break or free periods. The case study teacher also claimed that the head teacher sets up meetings between TDP and non-TDP teachers specifically to share knowledge. However, non-beneficiaries complained that the TDP teachers did not actively share knowledge with them, and that the non-TDP teachers had to request permission to view the videos included on the 'trainer in the pocket'.

The timing of the fieldwork did not allow us to visit cluster meetings in this LGEA. TDP teachers said that all teachers participated during the cluster meetings. However, TFs complained about a lack of interest in attending the training. They noted that TDP had not met teachers' expectations in respect of the allowances given to attend training (NGN 2,500), which are low compared to other training sessions. This allowance is allegedly insufficient to cover transport costs and to supplement the little refreshments provided at the training, which lasts for a whole day.

The cluster meetings are allegedly facilitated in a mixture of English and Hausa, since teachers would not understand if they were held just in English.

N.6 Katsina, low-performing school

Table 52 Some key quantitative baseline survey data for Katsina, low-performing school

State	Performance category	English TDNA (%)	Maths TDNA (%)	English literacy pupil test (%)	Numeracy pupil test (%)	# teachers	# pupils	Average class size
Katsina	Low	6	39	25	22	13	1010	68

Source: TDP in-service training impact evaluation baseline survey (2014); TDNA and pupil test scores represent marks scored (raw) as a percentage of total marks; # teachers = number of teachers employed (Grades 1–6); # pupils = number of Grade 1–6 pupils registered

N.6.1 Pupils

According to both the head teacher and the pupils themselves there is an issue with pupil absenteeism at the school. During the rainy season it was said that children work on farms, while in the dry season children – especially girls – hawk. Children in focus groups admitted to not coming to school sometimes. Girls talked about household chores, such as washing their own uniform and washing dishes, that could sometimes become too time-consuming and could result in them missing school. Boys all said that they went to work on the farm sometimes and sometimes missed school as a result.

However, children also said they liked coming to school and were happy there, when they were being taught. In the boys' FGD, the respondents said that they sometimes have two or three lessons in a day – they were not sure how many or what lessons they were supposed to have in a day. They spent some of their time sitting idly when not being taught, and they gave mixed messages about whether they liked this or preferred having the teacher in the class.

However, they described how children were sometimes beaten by teachers. Their example of a 'happy child' was a child who had a parent who would sometimes come to school to tell the teachers not to beat her. Children complained that the new head teacher of the school beat them for coming to school late.

One of the boys talked about his older sister who had stopped coming to school in Grade 3. Their mother had asked her to stop going to school and to sell *fura*. Her brothers and at least one of her sisters continued to come to school. The girl was due to be married soon.

N.6.2 Teachers

TDP teachers initially said subject knowledge was adequate and that the curriculum was too hard for the children. On being specifically asked if the problem was with the curriculum or with the teachers' subject mastery, they said that it was both. They said that teachers with an NCE are fine but there are problems in the state with teachers using an NCE certificate from someone else who has the same name, or even changing their names to get the certificate. These teachers, according to the TDP teachers' accounts, have problems with language. Teachers who have really gone through the teacher training will have enough knowledge to understand the curriculum. However, TDP teachers admitted finding it hard to teach the curriculum. They said that once you have an

NCE it is presumed that you can teach all subjects, whereas their knowledge was in fact limited to specific subjects.

The case study teacher in this school used actual money to teach the children about money. He showed money to the pupils and showed them money sums on the blackboard. The class initially worked all together, with individual pupils coming forward to write answers on the board. The teacher then divided the class into one group of boys and two of girls. However, only the seats and desks towards the front of the class were arranged into groups. He used some elements of group learning, although he focused on one group at a time, and when one group was being taught the others were simply waiting. He showed the use of counting for the addition of numbers and used Naira notes to talk about denominations of money. Throughout the lesson, the teacher had the cane in his hand, but did not use it; nor did he scold the children verbally.

Some teachers lived locally while others lived in a town around a 30-minute drive from the school. Teachers who came from the local community sometimes said they were motivated by a desire to see the community grow. Our case study teacher said he was motivated by his friendships with people living in the village, and by the fact that pupils were happy when he came into the class.

TDP teachers said that regular inspections were demotivating, and that the inspectors came to look for faults only. They said that the inspectors just looked at lesson plans and not at what the teacher was actually doing. They suggested that the inspectors should examine the pupils too, to understand how well they were being taught. However, the Education Secretary refuted this, saying that inspectors went to the school to inspect and encourage, rather than to find fault or place blame. Grievances over promotion were also raised by non-TDP teachers.

Teachers complained about infrastructure and class resources, saying there were not enough textbooks, that they wanted more teaching aids, and that there were too many children in some grades.

The teachers in this school did not do (or would not admit to) any additional work outside their teaching.

N.6.3 SLM

There were reportedly problems with teacher lateness, especially among the teachers who did not live locally. The head teacher also appeared to lack authority over at least some of the teachers. Some of the teachers would refuse to pay attention to the head teacher's attempts to correct them when they made mistakes, according to the head teacher and according to some of our teacher respondents.

The head teacher was a member of the local community and said that members of the community respected him a lot, and that he would advise parents and they took his advice – for example, on children who were absent from school.

Under a previous head teacher inspectors from the SUBEB had once visited the school and found the head teacher and several teachers absent: all of them were posted to more remote rural schools as a punishment. One of the teachers who was not absent was then made acting head teacher. More than a year later, he remains acting head teacher and has not been formally promoted to the role.

Some teachers complained that a lack of leadership from the head teacher was demotivating for them. They argued that the head teacher should have been appointed from another town, and that a head teacher promoted from within a school is not respected by the teachers.

The head teacher felt that the infrastructure within the school was not sufficient for the number of children: Grades 2 and 3 were split into two arms each but other classes had single arms and were, according to him, large. However, we only observed 23 children in the Grade 2 class, possibly reflecting pupil absence.

We observed that the school had a very well-equipped nursery compared to the other classrooms in the school. The government had apparently provided furniture for the nursery, while the nursery teacher himself had provided decoration and learning materials.

The Education Secretary complained of difficulty in punishing teachers because their payment comes from the SUBEB rather than from the local government.

N.6.4 Community and parental engagement

The head teacher of the school alleged that parents do not expect to get any benefit from sending children to school, and so do not send them. According to this account some young people who have finished school end up unemployed, spending their time idle in the village. According to the head teacher and teachers, there was good communication between them and the community. Teachers said the community respected them, and they were proud that some of the adults in the community had been taught by them. The case study teacher's photos were nearly all of scenes in the local community, suggesting quite a close link with his neighbours in the village.

N.6.5 TDP processes

TFs said that the challenges they faced included a reluctance by teachers to adapt to new methods, especially for older teachers. They believed that the teachers did not really put the new methods into practice because of this. They also said that other (non-TDP) teachers could undo any progress they have made among the children taught by TDP teachers, by continuing to use the older methods. Teachers with lots of experience, according to the facilitators, already know how to teach. They also claimed that the TDP teachers are not really sharing their knowledge.

This latter claim was endorsed by the non-TDP teachers, who said that they had not seen any of the TDP materials. They felt there was no problem with their teaching methods. By contrast, TDP teachers suggested there was a divide between the TDP and non-TDP teachers, and argued that non-TDP teachers did not want to see the materials. Like the facilitators, TDP teachers argued that the non-TDP teachers would undo changes they made – for example, to the arrangement of desks in the classroom (into groups versus whole-class teaching). The selection of three teachers for TDP was causing some tensions and discord within the school, possibly worsening the existing divides between teachers from the local community and those living in the town; and between those who sided with the head teacher and those who did not accept the head teacher's authority.

TDP teachers suggested that TDP should provide a certificate, and even that the TDP certificate should be a prerequisite for registration with the Teachers' Registration Council.

The TDP teachers felt that the cluster meetings were useful, because the attendees learned new skills and techniques, but they had not so far found the school support visits to be valuable. They claimed that the TFs came and observed their lessons – they appreciated their work if they had done well, but they did little else.

The TFs complained that schools in remote villages were hard to reach, and they were not provided with any transport.

The Education Secretary was clear that he had selected the teachers for TDP in that LGEA. One of the teachers on the original list had been sent away – posted to another school – and so had been replaced with a nursery school teacher. The Education Secretary said he had considered for inclusion in TDP those schools with at least six classes, where the teachers had an NCE, where they taught Grades 1–3, and where the teachers were not about to retire.

A nursery teacher was among those receiving the TDP training. As has been said, he had replaced the teacher originally selected, who had been posted to another school. The nursery teacher said that TDP helped to simplify his teaching, especially for nursery students.

N.6.6 Summary

Despite its rating as a 'low-performing' school, this school had better infrastructure than some other schools. Teachers were sufficient in number and present in the school but many attended late and were not teaching their classes. There was evidence of discord and of the head teacher lacking authority over some of the teachers. Troublingly, this discord appeared to have been worsened by the introduction of TDP, as some teachers resented not having been included in the programme. As in other schools, there were issues with children being absent to work on farms or in the household. Teachers and the head teacher seemed to have good links with the local community, but raised doubts about the extent to which parents in the community supported education.

N.7 Zamfara, high-performing school

Table 53 Some key quantitative baseline survey data for Zamfara, high-performing school

State	Performance category	English TDNA (%)	Maths TDNA (%)	English literacy pupil test (%)	Numeracy pupil test (%)	# teachers	# pupils	Average class size
Zamfara	High	30	50	18	42	27	1125	21

Source: TDP in-service training impact evaluation baseline survey (2014); TDNA and pupil test scores represent marks scored (raw) as a percentage of total marks; # teachers = number of teachers employed (Grades 1–6); # pupils = number of Grade 1–6 pupils registered

N.7.1 Pupils

Children carried out similar chores in the house to those carried out in a rural settings, including fetching water and washing plates. The children we spoke to in focus groups said that they did not have to work outside of the home, but the head teacher and assistant head teacher told us that some of the other children enrolled in the school do work on farms (despite the school's urban setting, there was apparently farmland nearby) or hawking goods. Parents especially tended to take their children out of school on a Friday, according to the assistant head teacher.

Children were able to recall the subject matter of their previous lessons, but struggled to recall the content. For example, the girls we talked to had studied parts of the computer the previous day, but could not tell us any of the parts.

When there was no teacher in the class (see below) we observed children sitting quietly in some cases, and playing or fighting noisily in other cases. We did not observe them studying in classes without a teacher; they did not have textbooks or exercise books on their desks.

A few of the boys in the school were resident in nearby almajiri schools; their parents, and the almajiri school head teacher, permitted them to come to the Western school.

N.7.2 Teachers

Teachers said that they were demotivated by not receiving their salaries on time, and according to the head teacher some had left the profession for this reason in the past. The level of the salary, as well as regularity of payments, was said to be a problem.

Unlike most rural schools, this school had a mixture of female and male teachers.

Despite some of the teachers starting to use TDP and Jolly Phonics methods, pedagogy in the other classes in the school appeared to rely heavily on repetitive chanting of set phrases or lists in English, with further explanation sometimes given in Hausa. Teachers felt that their training

provided them with the knowledge they needed to cover the curriculum, but that the problem was with children's backgrounds.

One teacher openly admitted to working outside the school as a carpenter and to having lied to the head teacher in the past so that he could go to his job. Some teachers also went to work on farms, although this apparently only happened after working hours.

N.7.3 SLM

The school has relatively good infrastructure, with more than one classroom for each grade, a library and an ICT block. However, the ICT block was empty and unused. According to the head teacher and assistant head teacher there were also insufficient classrooms for the number of pupils. During our visit, we observed both classrooms with very few students as well as quite overcrowded classrooms. Grades 1 and 2 had a single classroom, while Grades 3–6 were each split into two arms. Grade 2 apparently sometimes had more than 100 pupils, although at the time of our visit there were fewer, apparently due to pupil absenteeism. The two arms of Grade 3, on the other hand, both had very few children. Some of the classrooms had insufficient desks, so that children were sitting on the floor.

The school also had outward signs of being relatively well-organised and managed. For example, there were timetables in each classroom, and elected monitors in each class. If a teacher was not present, the monitor was supposed to check the timetable and request the appropriate teacher to come and teach them.

In practice, however, we nevertheless observed several classes without teachers, while at the same time a number of teachers sat outside the classrooms and talked. Some of these were student teachers, who appeared to spend a lot of time reading textbooks or writing (perhaps lesson plans or teacher journals). One teacher in the school was hardly trusted to teach at all, and was given the task of teaching 'some basic things in mathematics' to the Grade 1 children. The head teacher expressed frustration with the quality of teachers who, despite having passed an NCE or undergraduate degree in education, were not able to teach.

Teachers appeared annoyed when (in our presence) the assistant head teacher asked them to take a class. There were several student teachers, who appeared to be given few classes to teach, but spent some time observing the other teachers. Our pupil focus groups were conducted late in the morning, yet children had not received any lessons by the time of the focus group. The head teacher argued that a delay of almost two months in salary payment might be the reason that some teachers had not taken their timetabled classes.

In the past, teachers had been transferred into and out of the school on grounds of their political connections.

The head teacher occasionally gave small sums of his own money to teachers as incentives. Two teachers were working on a voluntary basis, but receiving occasional payments from the head teacher's own pocket.

N.7.4 Community and parental engagement

The SBMC included a person who was influential in local politics, who was able to mobilise resources for the school. The head teacher noted that this person could go directly to the SUBEB,

for example – something that he himself could not do. Even the prominent SBMC member was not guaranteed an audience with the head of the SUBEB, but he could at least approach him.

Most teachers felt that the community valued their work, although one said that community members knew that teachers would often not get paid, and as a result they would lose social esteem.

As elsewhere, small amounts of money were contributed by parents in the form of PTA fees, and were used for maintenance of the school.

N.7.5 TDP processes

It appeared that the TDP training received so far was causing some confusion in the school, with, for example, children and teachers mixing up sounds of letters with the names of the letters. Some teachers in the school had been trained in Jolly Phonics and it seemed that this might be causing some further confusion, given small differences in terminology and approach used between Jolly Phonics and TDP. Teachers told us that they mixed some of their former methods with the new methods from TDP: for example, continuing to give children notes to copy at the end of each class. They argued that TDP provided them with easier methods to make children understand quickly, unlike their previous methods in which they would invest a lot of effort but find that children still did not understand.

The head teacher had selected teachers for TDP. He said that TDP beneficiary and non-beneficiary teachers were sometimes asked to teach a class together, so that the non-beneficiaries could learn.

As elsewhere, there was some concern about delays in the payment of allowances for attending cluster meetings. The head teacher pointed out the transport costs that the teachers would face in travelling to the meetings meant that they would sometimes be out of pocket while waiting for their payments. The TF complained that resources were not adequate and that he had to provide his own materials for the cluster meetings.

Teachers argued that for TDP to work there needed to be more teaching materials, and they needed more supervision.

N.7.6 Summary

This school was relatively well-resourced, thanks in part to the strong connections it had with influential members of the community, and its urban location. It also had a large number of teachers present on the school grounds. The pupils in the school were not from rich backgrounds and some worked outside of, or during, school hours – although their parents are likely to have been richer in absolute terms than those in some of the remote rural schools we visited. The school also benefited from interventions by the British Council and Jolly Phonics, as well as TDP. Despite its relatively advantageous position, it continued to have evident problems with teacher absence from class, possibly linked to salary non-payment and a strained authority relationship between the head teacher and teachers. When in class, most teachers also continued to teach in a way that emphasised repetition and the copying of a limited number of set phrases and lists, and children showed limited signs of having learned anything in their recent lessons.

N.8 Zamfara, typical school

Main themes: Poor infrastructure, low pupil and teacher attendance, poor SLM, low teacher motivation (non-payment of salary), multi-grade teaching

Table 54 Some key quantitative baseline survey data for Zamfara, typical school

State	Performance category	English TDNA (%)	Maths TDNA (%)	English literacy pupil test (%)	Numeracy pupil test (%)	# teachers	# pupils	Average class size
Zamfara	Typical	23	32	13	25	5	256	65

Source: TDP in-service training impact evaluation baseline survey (2014); TDNA and pupil test scores represent marks scored (raw) as a percentage of total marks; # teachers = number of teachers employed (Grades 1–6); # pupils = number of Grade 1–6 pupils registered

This school is a small school, located in a slightly interior location, around 30 minutes' drive from Gusau, the capital of Zamfara state. It was established in 1979 as a school with two classrooms and since then there had been no addition to the school's infrastructure, even though pupil enrolment had multiplied quickly. The school has a third room which is used as the head teacher's office, as a staff room, as a library, as the school storeroom, and as a sitting room for visitors. One of the two classrooms, according to the head teacher, is almost unusable due to a leaky roof and rough muddy floors with large holes – as a result all pupils from ECCE to Grade 6 sit in a single classroom and are taught together. In the dry season classes can sometimes be held outside under a tree, but in the rainy season classes cannot be held outside or indoors since the roofs leak in both classrooms. Thus, classes are either suspended or held in the community head's home.

N.8.1 Pupils

The school had about 250 enrolled pupils from ECCE to Grade 6, but on none of the three days of the survey were there more than 50 pupils attending school. One of the reasons mentioned by the head teacher was that pupils were away on farms since it was harvest season (June). However, another reason for poor pupil absenteeism was that one of the three communities which constitutes the school's catchment area had decided to withdraw their children from the school because the parents from the community felt they were not learning much in this school.

During FGDs with pupils notebooks showed that the last lessons were held at least a couple of weeks ago. Pupils' writing in notebooks, even though checked and signed by teachers, appeared almost illegible.

N.8.2 Teachers

Teachers at this school had not been paid their salaries for two months, at the time of the survey in June 2015, which expectedly had led their morale and motivation to hit a low point. Late payment of salaries, according to the teachers, had weakened their intrinsic motivation to teach and to do well as teachers, a role they claimed they found 'very interesting'. They often had to struggle to manage money to fuel their bikes to come to school or pay for public transport.

Two of the four TDP teachers mentioned their initial motivation to pursue careers in teaching being driven by 'the good you can do' for the life and career of your pupils. The third teacher, however, had a remarkable story regarding why he became a teacher: his uncle used to be a teacher and was seemingly well-connected. This teacher from this school at that time was not employed, apparently just 'at the street playing ball'. His teacher-uncle asked him if he could teach, to which he replied in jest that he could ('it was like a joke'). That evening his uncle sent a messenger to his house with a letter of appointment to be teacher.

N.8.3 SLM

The head teacher of this school has been a lifelong classroom teacher who was posted to this school in 2014. In this school, aside from being the head teacher, he also taught English, Hausa and art and Craft.

The head teacher expressed a common complain of not having any authority in the recruitment of teachers to his school, and complained about how teachers with no apparent pedagogical skills were being posted to the school just because of their political connections: 'some people have Godfathers who stand for them'.

The head teacher appeared to have a lot of administrative, management and leadership responsibilities. It is therefore possible that his day-to-day responsibilities (e.g. ringing the bell, applying first aid to injured children, maintaining records, rationing chalk, etc.) leave very little mental bandwidth to think about longer-term SLM issues.

The head teacher claimed that he conducted frequent lesson observations and often had to correct spelling mistakes made by teachers on the blackboard – to minimise any embarrassment for the teacher he claimed to often go and silently whisper the correct spelling to the teacher ('emergency correction').

The local community and SBMC, the head teacher claimed, respected him and regularly offered him food and small gifts.

The head teacher listed a number issues which he found to be demotivating: the lack of teaching aides, especially textbooks (which he often had to purchase himself); the lack of classrooms; delayed promotion; and late salary payments.

N.8.4 Community and parental engagement

One of the three communities which constitutes the school's catchment area had decided to withdraw their children from the school because the parents from the community felt the children were not learning much in this school, and instead would make better use of their time on the farm.

This school has historically had frequent transfers of head teachers and teachers, which led the local communities to grant plots of land adjacent to the school to the head teacher and teachers of the school. The teachers can farm on the land and use the produce for their own income/consumption. They have to give up their right to the land when they leave the school. This, the head teacher claimed, was the communities' effort to retain good teachers in the school and to build their attachment towards the school and communities. One downside to this was that

the head teacher and teachers often spent time on the farm during school hours, and also engaged pupils to help them with farming during lesson time.

The head teacher had recently convinced community members to open a Community Social Development Programme (CSDP) account in the name of the community. Under this scheme the community would contribute 10% of the money required to build an additional block of classrooms in the school while CSDP would contribute the rest.

N.8.5 TDP processes

Of the four teachers (including the head teacher) selected for the programme from this school, two of the teachers were absent throughout the three days of the survey team's stay in the school. One of them mentioned he had no money left (after delayed salary payments) to fuel his motorbike to come to school. The other teacher was away on study leave.

The head teacher mentioned that the non-TDP teachers in this school were somewhat disgruntled about their exclusion from the programme and believed the head teacher had a role to play in it.

The head teacher also noted that without timely salary payments it is difficult for the teachers to practise some of the pedagogical techniques being imparted by TDP. For example, 'that is salary delay which makes a teacher not to have flash card for teaching or carbon sheet. Only when you get salary, you can buy them.'

In terms of usage of materials, the amplifiers given to this school did not appear to be in use and a pupil was sent to the head teacher's house to fetch the amplifier upon the arrival of the survey team. Similarly, printed materials, like teacher guides, given to the TDP teachers in this school lay almost wholly unused in the school library.

The head teacher and teachers in this school also complained in detail about the lunch they received at the cluster meeting: 'The kind of lunch they give us during cluster meetings "Wallahi" [I swear] I cannot give my family such kind of a food, very bad indeed, virtually all of us complained about the nature of the food.'

N.9 Zamfara, low-performing school

Main themes: Pupil absenteeism, teacher absenteeism, gender, strong head teacher leadership, salary payments, political appointments

Table 55 Some key quantitative baseline survey data for Zamfara, low-performing school

State	Performance category	English TDNA (%)	Maths TDNA (%)	English literacy pupil test (%)	Numeracy pupil test (%)	# teachers	# pupils	Average class size
Zamfara	Low	12	33	21	32	9	299	35

Source: TDP in-service training impact evaluation baseline survey (2014); TDNA and pupil test scores represent marks scored (raw) as a percentage of total marks; #Teachers = number of teachers employed (grade 1-6); #Pupils = number of grade 1-6 pupils registered

This school is a small primary school (350 pupils in Grades 1–6, of whom 79% are boys) established in 1972. The school is located in a rural farming village and has poor road access. The school has five classrooms for seven classes, including ECE, with average class sizes of 54 pupil and a moderate PTR of 42:1. The head teacher described the school a having better infrastructure, more staff and more pupils than other schools nearby. However, there are still insufficient classrooms, meaning that the ECE class must be taught under the shade of a tree outside.

UNICEF has built toilets and a borehole in the school. The borehole is the best in the village, and therefore attracts a large number of villagers and livestock, who visit the school grounds to obtain water, which sometimes disrupts pupils and teachers in classes.

N.9.1 Pupils

Pupils in this school were frequently out of the classroom, though not to the same extent as in other rural schools observed as part of the qualitative study. One potential indicator of the extent of local poverty was the scarcity of school uniforms compared to other schools. As a result, it was sometimes difficult to distinguish between students and out-of-school children who were in the school to collect water. During the FGD pupils revealed that they wished that they could create a new uniform and ensure that all students were able to attend in this new uniform.

Pupil absenteeism is a major problem in this school, driven in part by low girls' enrolment and farming imposing an additional obligation on children's time. Pupil attendance dropped even further following rainfall on one of the days the qualitative research took place. Some pupils also attend Arabic school in the morning or afternoon, in exchange for approximately NGN20 per month. Parents apparently often complain that their children do not go to school even when they are sent. TDP teachers ranked pupil absenteeism as the second-highest obstacle to teaching effectively.

Students in this school were unable to distinguish between the subjects (i.e. English, maths etc.) they had been taught. However, some did seem to have an awareness of the topics recent lessons had covered. For example, they knew they had been taught the sounds of different letters, but not that this was considered to be 'English'. Despite this, pupils could not read the writing they had copied in their exercise books and their handwriting frequently failed to distinguish words or even letters.

N.9.2 Teachers

There are nine male teachers and no female teachers in this school. Some teachers at the school receive Jolly Phonics and TDP training. Training from the SUBEB and LGEA did not appear to be delivered at frequent intervals: one teacher reported receiving only two in-service training sessions in his 11 years as a teacher: the MDG workshop and another workshop organised by the SUBEB.

Jolly Phonics was providing flashcards for teachers to use in class. The case study teacher demonstrated some 'child-centred' teaching techniques:[50] for example, he used flash cards and a page torn from a textbook as teaching aids, grouped pupils for activities, and asked pupils to

[50] Note that, due to timetabling issues, the case study teacher had been inadvertently warned of the lesson observation and so had time to prepare an 'ideal' lesson.

participate in the lesson by getting them to write one-word answers to questions on the board. The case study teacher appeared to have tried to build flexibility into the lesson to allow pupils to express their creativity. For example, in one instance he told the pupils 'to write anything they feel like writing before the lesson starts, then [he] was going round to check what they [had] written'. This exercise was hindered by a lack of desks and writing materials.

Despite this, pupils' exercise books contained mainly text that they had been asked to copy from the board, and not independent work or answers to questions. Pupils did not seem to understand the concept of homework when asked about it. There also seemed to be a significant repetition of lesson content, with the case study teacher taking photos of the same lesson content that had been observed earlier in the week. Neither TDP nor non-TDP teachers believed that poor teaching technique was a major obstacle to effective teaching in the school.

Both groups of teachers believed that low teacher motivation was having a major negative impact on their ability to teach effectively. They do not have a close support network with teachers from other schools nearby. The main factor that demotivated teachers was late payment of their salaries. Teachers that live far away from school find it difficult to get to school under such circumstances, due to the cost of commuting, leading to unattended classes. They also have difficulty feeding their family under these conditions. Not only were teachers not paid, but their promotions have often not resulted in pay increases or their overdue promotions have been denied. They advised that teachers should not enter the career for the prospect of financial reward, but for 'the blessings of being a teacher'. These delayed payments encourage teachers to cultivate alternative income streams, such as farming. One teacher was so demotivated by the poor salary that he said he will leave his career in teaching as soon as he gets another job.

TDP teachers claimed that they had no obligations outside of school that prevent them from carrying out their jobs, except in the case of illness or meetings at the LGEA. However, the non-TDP teachers admitted to having other sources of income when probed. One teacher has a barber's salon, rents canopies and has a farm. Another has a farm, repairs motorcycles, and does manual labour. The third also has a farm and teaches lessons in another school in the afternoon. Each of these reported that they would only carry out these other roles outside of the school hours, claiming that their 'business does not affect [their] teaching job, but the teaching job affects business'. However, the case study teacher privately confessed that 'nobody questions your coming late to school: some teachers will come to school at 9am or 10am and they will record 8am in the time book'. It is possible that non-TDP teachers spoke more freely of their other income-generating opportunities because they did not fear having TDP training opportunities withdrawn in the same way as TDP beneficiaries did. Pupils reported teacher absenteeism as one factor that makes happy children sad at school, and some students claimed that they had not been taught English for more than two weeks. The head teacher would allegedly sometimes authorise leave for teachers to tend to their farms.

The head teacher complained that fresh graduates from teaching college do not have sufficient subject knowledge to teach effectively. However, TDP teachers did not recognise poor subject knowledge as a constraint to teaching effectively, leading them to rank it as the lowest of the 12 possible obstacles. However, poor teacher subject knowledge was clearly a constraint, as reflected in various serious spelling errors that often made writing on the board nonsensical (e.g. 'sup ingridings' instead of 'soup ingredients').

N.9.3 SLM

The current head teacher was appointed as head teacher in 2013, having previously been head teacher and assistant head teacher at different schools nearby. He demonstrated a generally pro-active approach to SLM and is well-respected by his teachers. For example, his office and the staff room was well-organised (though sparse, given the lack of textbooks): the head teacher had made and displayed posters summarising some of the key 'learning points' from the Jolly Phonics and TDP training on the walls of the staff room. The head teacher encourages the use of improvised teaching aids, and showed an example in the form of weighing scales made using two guinea corn stalks and a nail.

The head teacher identifies 'successful' teachers through lesson observations. Teachers claimed that they were motivated by lesson observations and the feedback they receive, which they feel helps them to improve their teaching technique. However, the head teacher asks teachers about their pupils' performance only every two to four weeks. In addition to advising on teaching technique and correcting mistakes in subject knowledge, he challenges teachers that curtail their lesson time. Where possible, he rewards teachers that perform well with soya bean cakes bought out of his own salary, or by arranging transport back to the village for them. The head teacher reports teachers that are absent from school to the Area Development Officer. He feels he has no authority to discipline teachers himself, but can threaten to report them. Education officers from the LGEA, SUBEB and TDP inspect the school and leave feedback in the visitors book. However, the extent of this feedback is limited to 'good', 'satisfactory', or similarly high-level comments.

The head teacher is not satisfied with how the recruitment process works due to the interference of political influence leading to unqualified teachers being posted to the school. He has no influence over which teachers are posted to his school and has difficulty in retaining quality staff. In contrast, the head teacher complained that urban schools receive many more teachers than they need for their size. For example, he knows of one school that receives 30 teachers in the morning and afternoon. He said that these teachers tend to use their political connections to be posted to urban schools, which means that teachers are not allocated to schools where they are needed most. In some schools in the local town, 15 teachers are apparently responsible for teaching just Grades 1–3.

One example of political influence in the recruitment process was provided by a teacher. Her brother, who works in the LGEA, had secured her an offer to be a teacher in 2008, when she was still in school studying. During this period, the brother collected the salary and passed a percentage on to the teacher, despite neither of them actually teaching. Once she completed her studies in 2010 she took up the full-time teaching job and received the salary herself in full. Some others allegedly held up to five offer letters at once, meaning that they collected the salaries of five teachers.

The head teacher was severely demotivated by delayed salary payments, which had not been disbursed since April (for three months). He was also demotivated by the lack of sufficient textbooks, which he now has to borrow from head teachers at other schools with which he has a good relationship. The PTA is currently trying to raise funds to buy textbook for the pupils.

N.9.4 Community and parental engagement

The head teacher felt that the community values his work, as evidenced by an alleged increase in pupil enrolment over the past two years. However, the TDP teachers felt that only the more

educated members of the local community value the work they do as teachers. For example, they blamed the high pupil absenteeism during the planting and harvest seasons on lack of parental support.

The head teacher maintains a positive relationship with the SBMC, who had previously raised funds to build the bridge that leads to the local community to give road access, and money for repairs of the windows and doors in the school. There is a community training programme for members of the SBMC, including the head teacher, head boy, and leaders from the local community, to help them carry out their responsibilities effectively. The SBMC inspects the school. The TDP teachers also said that the SBMC played an active role in bringing together parents, teachers, and community leaders. However, the non-TDP teachers claimed that the last SBMC meeting was held five months ago.

One of the district elders is a former Arabic teacher at this school and a current SBMC member. He has apparently played a key role in encouraging parents to send their children to school. Teachers turn to the district head for support when they feel demotivated. The district head sometimes supports the teachers financially, given their late salary payments.

N.9.5 TDP processes

The head teacher has taken an enthusiastic attitude to encouraging teachers in this school to adopt the new teaching techniques promoted by TDP. In addition to the posters mentioned above, he has scheduled a formal meeting organised for both the TDP and non-TDP teachers to share knowledge and address challenges they face. However, the TDP teachers said that they had not shared knowledge with non-TDP teachers, despite being told to do so at cluster meetings.

The head teacher did not seem to be aware of how the TDP teachers were selected to benefit from the programme, despite this being his formal responsibility under the TDP design. This suggests that beneficiaries were selected by someone other than the head teacher.

The amplifier in the school was not working because of a problem with the batteries, so they were not using this. Poor network coverage and flat batteries made it more difficult to coordinate TDP teachers in rural areas. In Zamfara, TDP phones were also pre-loaded with the Jolly Phonics training materials.

The TFs in the LGA were a teacher in an urban model school/staff in the LGEA and an employee of the Science Technical College Board. The TFs divide the 12 schools they are responsible for between them, making them each responsible for six schools. They regarded their primary objective as being to train teachers to adopt child-centred learning techniques.

Cluster meetings were reportedly held every month; however, beneficiaries claimed that there was one month when the meeting was missed. They were enthusiastic about the cluster meetings as the distinguishing feature of TDP, since they believe that they have created a close relationship among teachers in the LGA and paid special attention to individual teachers by giving them phones. Despite this praise, the case study teacher could not remember much that was covered during the cluster meeting held previously that week, suggesting that the learning had not been reinforced. There was some confusion as to whether TDP was teaching subject knowledge or pedagogy: the TFs claimed that 'Jolly Phonic teaches content while TDP teaches technique'.

Teacher turnout at cluster meetings was estimated at about 70%. The TFs believed that the lack of adequate payment of transport allowances was a key driver behind late-coming and teacher absenteeism, especially when combined with the delayed payment of salaries. Teachers also suggested that TDP should pay transport allowances promptly, because transportation challenges posed a significant obstacle to attendance at the cluster meetings for some. Where teachers have complained to TDP about this, they have been asked to provide a bank statement as evidence that they have not yet received payment, which has led to frustration with the programme.

Insecurity was also mentioned by one attendee at the cluster meeting as a barrier to attendance to those coming from more rural districts (not this school) in the LGA affected by communal violence. This insecurity apparently makes it unsafe to travel on the roads from late afternoon. Given the fact that the cluster meetings finish at 4.30pm, this means that teachers would prefer to stay the night in the town where the cluster meeting takes place. However, this is expensive and therefore is a disincentive for teachers to attend.