

Teacher Development Programme (TDP)

Evaluation Framework and Plan

Final version

EDOREN

10 December 2014

Acknowledgements

This draft evaluation framework was developed by EDOREN, and specifically Sourovi De, Ifeatu Nnodu, Alex Hurrell, David Megill, and Ian MacAuslan. Thanks are due to various EDOREN and OPM colleagues for support and advice, including Matthew Powell, Tom Pellens, Stuart Cameron, Rachel Outhred, Vegard Iversen, Femi Adegoke, Oladele Akogun, Florian Friedrich and others.

Thanks are also due to the DFID Nigeria education and the TDP teams for extensive interactions during the course of developing the framework, and the review of initial drafts of parts of the document. In particular, we are very grateful to Bukola Oyinloye, Bob McCormick and several other colleagues at TDP, and to Esohe Eigbiki, Emily Oldmeadow and Stuart Martindale at DFID, for extensive comments and discussions around earlier drafts of this document.

The document has also benefitted from extensive peer reviews by Paud Murphy and Tom Pellens, for which many thanks. We are also grateful for a review from an anonymous SEQAS reviewer. All errors remain the authors'.

This assessment is being carried out by EDOREN (Education Data, Research and Evaluation in Nigeria).

The EDOREN Country Director is Oladele Akogun at Oladele.akogun@edoren.org. The project manager is Ian MacAuslan at ian.macauslan@edoren.org. The contact point for DFID is Emily Oldmeadow at Emily.oldmeadow@dfid.gov.uk.

Disclaimer

EDOREN is a consortium of leading organisations in international development and education: Oxford Policy Management (OPM), and the Institute of Development Studies (IDS), at the University of Sussex, and is supported by UK Aid. EDOREN cannot be held responsible for errors or any consequences arising from the use of information contained in this report. Any views and opinions expressed do not necessarily reflect those of OPM, IDS and EDOREN or any other contributing organisation.

EDOREN	8B Winnipeg Close	Tel	+234 803 722 0460
Education Data, Research & Evaluation	off Panama Crescent	Email	office@edoren.org
in Nigeria	Maitama District	Website	www.edoren.org
	Abuja, Nigeria		

Executive Summary

S.1. This document sets out the proposed framework and plan for the evaluation of the Teacher Development Programme (TDP). Funded by the UK Department for International Development (DFID) and implemented by Mott MacDonald, the TDP seeks to improve primary and junior secondary learning outcomes in six states of northern Nigeria by improving the quality of teachers through in-service and pre-service training in two phases. Working in Jigawa, Katsina, Zamfara (phase 1, 2013-2019) and Kaduna, Kano and Niger (phase 2, 2016-2019), the TDP overlaps with other DFID-funded education programmes that also seek to improve learning outcomes: the Girls Education Project (GEP3) that works in Katsina, Zamfara and Niger, and the Education Sector Support Programme in Nigeria (ESSPIN) that works in Jigawa, Kaduna and Kano. More information on the TDP is provided in section 1, which discusses the TDP's theory of change in detail, and Annex B, which provides detail on specific interventions.

S.2. The main objective of TDP is to improve the skills of teachers in the three core curriculum subjects of English, maths, and science & technology and produce better teachers through a combination of in-service and pre-service interventions. The specific objectives of TDP are therefore to: (a) improve in-service training of primary and junior secondary school teachers (this is the major component of the TDP); (b) develop more effective teacher educators of primary and junior secondary school student-teachers through system reform; and (c) strengthen evidence-based research to influence and inform decisions on teachers' effectiveness and efficiency. These are discussed below in detail in Section 1.2.

S.3. This evaluation framework has been designed by Education Data, Research and Evaluation in Nigeria (EDOREN), a DFID-funded education research and evaluation project in Nigeria managed by Oxford Policy Management (OPM). The terms of reference (TOR) for the framework design are in Annex A, though the scope of the evaluation framework was later revised only to include an evaluation of the in-service component of the TDP.

S.4. Given the gaps in research and policy understanding, and the TDP's balance of spending, the key question the evaluation is expected to answer is whether and how TDP's in-service teacher training model improves teacher quality and subsequently learning outcomes of pupils in primary education in northern Nigeria. This will fill key gaps in global understanding of what works to improve teacher quality in resource poor contexts. This is expected to be useful for DFID in Nigeria and globally, and for the state and national governments in Nigeria, and for those interested in improving teacher quality in other resource poor contexts. In addition, the evaluation should provide formative information to help TDP design and implement its second phase, which starts in 2016, and evidence on synergies between the in-service and pre-service components, without directly evaluating the pre-service component. Section 1.3 discusses the purpose of the evaluation, and its intended audience, timing, resources available and the evaluation's fit with TDP's own results and evidence output. The role of other stakeholders in the evaluation is explored in more detail in section 6.

S.5. In order to do this, this evaluation framework takes a theory-based approach to answer questions about the relevance, efficiency, effectiveness, impact and sustainability of the TDP in-service output, following standard evaluation criteria and principles. These background principles are discussed in section C.1, and detailed questions derived from the key purposes set out above are provided in Annex G. These specific questions were developed in consultation with DFID and TDP, and from a reading of TDP's theory of change and implementation strategy, and the wider evidence base.

S.6. The evaluation plan proposed in this document focuses on phase 1 of the TDP in-service programme. The key questions are summarised in the table below.

Table 1: Key TDP evaluation questions		
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

S.7. Given the focus on understanding the impact of TDP's in-service component, and the need to distinguish from other education programmes, the centrepiece of the evaluation is an experimental study design based on constrained randomisation that will allow us to measure and attribute the quantitative impact of TDP's in-service work in phase 1 on learning outcomes and teacher quality. This requires that TDP, and particularly TDP's in-service state team leads, work closely with the State Education Boards (SUBEBs) in intervention states to ensure that the intervention and control schools are selected in a way that allows a randomised assignment of schools to intervention and control clusters. Specifically, the SUBEBs in each state need to select two clusters of 12 schools in each of the 14 pilot Local Government Education Area (LGEA), and the evaluation team would then randomly allocate treatment to one cluster and the other, i.e. set of 12 schools in the control cluster, would not receive TDP in-service interventions. This procedure, agreed by TDP and DFID, is the only way that the evaluation can confidently attribute changes in learning outcomes and teacher quality to TDP's in-service activities in each state, which is a priority for DFID. However, it requires TDP to work closely with the SUBEBs to ensure that they understand and accept this process. If this quantitative attribution evidence were not required or this procedure proved unworkable, the evaluation design would change significantly from what is proposed here, to put much more emphasis on generating qualitative data to assess TDP's contribution to change along the results. However, it can be reported that by early November 2014, this school selection process has largely progressed as plan, and the EDOREN in-service baseline evaluation teams had received cluster lists for Jigawa, Katsina and Zamfara and random assignment of clusters to treatment and control clusters had been completed in all three states.

S.8. These survey-based methods will be supported by primary qualitative fieldwork and secondary data collection designed to test TDP's theory of change and to provide responses to other evaluation questions (particularly around relevance, efficiency and sustainability), and to assess synergies with the pre-service component. Section 4 describes the approaches and methods proposed, including details of randomisation, sample sizes and implementation modalities, and C.1 and Annex F provide technical detail.

S.9. The evaluation plan in section 4 includes detailed timings for the evaluation fieldwork and results, particularly for the programme's in-service component. Since TDP is divided into two phases, and the evaluation of in-service phase 1 needs to provide formative information on in-service phase 2 (to be implemented in 2016), these timings are sensitive, and the quantitative fieldwork in particular needs to be fixed early on. The following schedule is proposed, with three key surveys:

Table 2: TDP in-service evaluation quantitative fieldwork timing			
Survey timing	Location	Phase	Purpose: evaluation of the impact of in-service training on:
October 2014	Phase 1 states (Jigawa, Katsina and Zamfara)	Baseline	Teacher quality and learning achievement in English, maths and science & technology for TDP-selected teachers and a panel of pupils starting p3 (tested on the p2 curriculum)
October 2015	Phase 1 states (Jigawa, Katsina and Zamfara)	Midline	Teacher quality in English, maths and science & technology for TDP-selected teachers, to provide formative information to feed into TDP phase 2 design.
October 2017	Phase 1 states (Jigawa, Katsina and Zamfara)	Endline	Teacher quality and learning achievement in English, maths and science & technology for TDP-selected teachers and the same panel of pupils now starting p6 (tested on the p5 curriculum) to provide summative evaluation.

S.10. The expected resources for these surveys will be: GBP700,000 for the baseline, GBP600,000 for the midline, and GBP700,000 for the endline, and the qualitative fieldwork envisaged would cost GBP390,000, making the total budget for the evaluation GBP2,390,000.

S.11. This proposal is broadly in line with the TDP logframe, and extends TDP's Research and Evaluation Strategy (McCormick 2013), although the constrained randomized evaluation approach is not envisaged in this strategy. However, one implication of the current evaluation proposal is that the evaluation does not generate an estimate of the impact of phase 2 of the in-service component, as envisaged by the logical framework. This is because existing evidence suggests that it is unlikely that learning outcomes will change in a short period and because it would be more expensive to measure learning outcomes as well as teacher effectiveness. In addition, the current evaluation design does not propose to test students at p4 level, because it is taking a cohort-based approach (i.e. following a group of pupils from p3 through p6, testing them on p2 to p5 curricula respectively) and this is more cost-effective to assess impact. The evaluation will also draw on the sources set out in the logframe to answer some of the questions above.

S.12. The introduction provides the background to the TDP and describes the programme's activities and theory of change. Section 2 outlines the purpose of the evaluation. Section 3 details the evaluation framework, including its key users and their needs, resources and timing, and the TDP's own results and evaluation strategy, and the evaluation framework, including criteria for what to evaluate, evaluation questions, and indicator areas. Section 4 sets out the evaluation plan for a theory based evaluation. Section 5 describes the evaluation's governance and management arrangements, including how we operationalise research ethics and independence, and comments on risks of Ebola and the elections in February 2015. Section 6 outlines an overview strategy for stakeholder engagement and communication.

Table of Contents

Acknowledgements	i
Executive Summary	ii
List of Tables and Figures	viii
List of Abbreviations	ix
1 Introduction	1
1.1 Background	1
1.1.1 TDP's role within DFID Nigeria's portfolio of education programmes	1
1.1.2 State of the Nigerian school teacher	2
1.1.3 State of Nigeria's colleges of education, teacher-educators and student-teachers	4
1.2 TDP's objectives and outputs	4
1.2.1 Programme objectives	4
1.2.2 Programme outputs	5
1.3 Evaluation purpose	7
1.3.1 Learning from the TDP for other contexts	7
1.3.1 Formative evaluation to help the TDP learn	8
1.3.2 Evaluation for accountability of TDP's performance	8
2 The TDP's theory of change	9
2.1 Unpacking the TDP's in-service theory of change	10
2.1.1 TDP in-service strategies	10
2.1.2 TDP strategies within a theory of change	11
2.2 Conclusion to the theory of change	25
3 Evaluation design and framework	26
3.1 Evaluation users	26
3.2 Evaluation resources	26
3.3 Evaluation timing	27
3.4 The TDP's Research and Evaluation (R&E) Strategy and logical framework	28
3.5 DFID five key criteria for deciding what to evaluate	29
3.6 Evaluation questions	30
3.7 Theory based evaluation	31
3.7.1 Map out the TDP's causal chain	32
3.7.2 Understand the TDP's context	32
3.7.3 Is the TDP going to have heterogeneous impact?	32
3.7.4 Can we establish the TDP's impact using a credible counterfactual?	33
3.7.5 Rigorous factual analysis across the TDP's causal chain	33
3.7.6 Use mixed methods	33
4 Evaluation plan: proposed fieldwork	36
4.1 Rigorous impact evaluation design	36
4.1.1 Research questions	36
4.1.2 Measuring learning outcomes	37
4.1.3 Overall approach	37
4.1.4 Timing of survey rounds	40
4.1.5 Instruments	40
4.1.6 Survey sampling design	41

4.1.7	Internal validity	46
4.1.8	External validity	48
4.1.9	Phase 2 TDP states: Kano, Kaduna and Niger	48
4.2	Rigorous factual analysis	48
4.2.1	Qualitative methods	49
4.2.2	Proposed qualitative fieldwork	51
4.2.3	Limitations	60
4.2.4	Process evaluation	61
5	Evaluation governance, management, independence and ethics	62
5.1	Governance and management	62
5.2	Independence, credibility and usefulness	63
5.3	Research ethics	64
5.4	Risks	64
6	Strategy for stakeholder engagement and communication	66
6.1	Role and involvement of stakeholders	66
6.2	Stakeholder involvement in framework development	67
6.3	Strategies to communicate findings	67
References		69
Annex A	Terms of reference	71
A.1	Summary	71
A.2	Background	71
A.3	The Objective	73
A.4	Recipient	73
A.5	Scope of the Task	73
A.6	Deliverables	76
A.7	Timeframe	76
A.8	Proposed skills mix of the team	76
A.9	Coordination and Logistics	77
A.10	Reference Documents	77
Annex B	Intervention factsheets	79
B.1	Output 1: In-service training component	79
B.2	Output 2: Pre-service training component	82
B.3	Output 3: Results & Evidence (R&E) component	86
Annex C	The TDP logframe (version October 2014) and research and evidence strategy	90
C.1	TDP Research and evidence strategy	96
C.1.1	Monitoring	96
C.1.2	Evaluation	96
C.1.3	Instruments and logframe	97
Annex D	Principles, standards and criteria guiding TDP's evaluation	100
D.1	DAC evaluation principles	100
D.1.1	Impartiality and Independence	100
D.1.2	Credibility	100
D.1.3	Usefulness	101
D.1.4	Participation of donors and recipients	101
D.2	DFID Evaluation standards	101
D.2.1	Quality Standards	101
D.2.2	Ethics	101

D.3	DAC Evaluation criteria	103
D.3.1	Effectiveness	103
D.3.2	Efficiency	103
D.3.3	Relevance	103
D.3.4	Impact	104
D.3.5	Sustainability	104
Annex E	Description of Methodology for Calculating the Minimum Detectable Effect (MDE) for a Difference of Differences Estimate	105
Annex F	School selection guidelines to SUBEBS (Jigawa version)	109
6.4	Step 1: Select four similar clusters of schools in each LGA	109
6.5	Step 2: Select teachers for TDP training in the selected schools	109
6.6	Step 3: Assign two clusters per LGA to the programme	109
Annex G	Evaluation matrix	110

List of Tables and Figures

Table 1: Key TDP evaluation questions	iii
Table 2: TDP in-service evaluation quantitative fieldwork timing	iv
Table 3: Assumptions for TDP achieving scale in TDP schools	22
Table 4: Assumptions for TDP achieving scale in TDP states	23
Table 5: TDP in-service evaluation quantitative fieldwork timing	28
Table 6: DFID Five Key Criteria for deciding what to evaluate	29
Table 7: Key TDP evaluation questions	30
Table 7 Sample design summary	43
Table 8 Sample sizes (by survey round)	43
Table 9 Minimum detectable effects: state-level impact indicators	44
Table 10 Minimum detectable effects: state average impact across all three TDP states	46
Table 11. Qualitative question areas, by respondent and year	52
Table 12. Qualitative sampling	60
Table 13. Possible limitations of qualitative component	61
Figure 1 TDP's theory of change	10
Figure 2: TDP theory of change from activities to impact, scale and sustainability	13
Figure 3: TDP intermediate impact to final impact	14
Figure 4: TDP outcomes to intermediate impact	16
Figure 5: How the TDP aims to achieve scale and sustainability	21
Figure 6: TDP evaluation governance arrangements	62

List of Abbreviations

3ie	International Initiative for Impact Evaluation
ASC	Annual School Census
CoE	College of Education
DAC	Development Assistance Committee
DFID	Department for International Development
EDOREN	Education Operational Research and Evaluation Nigeria
EIA	English in Action
ESSPIN	Education Sector Support Programme in Nigeria
GEP3	Girls Education Project phase 3
JSS	Junior Secondary School
LGA	Local government area
LGEA	Local Government education area
M&E	Monitoring and evaluation
MDE	Minimum Detectable Effect
MLA	Monitoring learning achievement
NCCE	National Commission for Colleges of education
NCE	National Certificate of Education
NEDS	Nigeria Education Data Survey
NERDC	Nigerian Educational Research and Development Council
NGN	Nigerian Naira
NTI	Nigerian Teachers Institute
OECD	Organisation for economic cooperation and development
OPM	Oxford Policy Management
PTA	Parent teacher association
R&E	Research and evaluation
SAVI	State Accountability and Voice Initiative
SBMC	School based management committee

SPARC	State Partnership for Accountability, Responsiveness, and Capability
SSO	School support officer
TDNA	Teacher Development Needs Assessment
TDP	Teacher Development Programme
TRCN	Teachers' registration council of Nigeria
UBE	Universal Basic Education
UNICEF	United Nations Children's Fund
WDR	World Development Report

1 Introduction

In this section we first provide details of the background to TDP and the purpose of the evaluation. There is strong evidence and recognition that children in Nigerian schools are not learning and that the quality of teaching is a serious concern. This provides a compelling rationale for intervening in the domain of teacher education, training and development. However, the issues and constraints surrounding teacher effectiveness and quality are not limited to lack of teacher education alone, and thus need to be appraised in light of the broader issues within teacher management—remuneration and other incentives, deployment, professional progression and career path, mentoring and supervision, role of the head teacher, availability of materials in the classroom, etc. This myriad system of issues within teacher management and its interaction with teacher training and development, in particular, means that TDP’s impact on teacher effectiveness and pupil learning achievement will ultimately depend on a number of interacting factors and thus uncertainty bears on the likelihood of TDP’s purported positive impact on learning outcomes. We also introduce the objectives of TDP and activities within its three outputs: in-service training, pre-service training, and research & evidence. The evaluation framework and evaluation will focus, at DFID’s request, on the in-service training component.

1.1 Background

1.1.1 TDP’s role within DFID Nigeria’s portfolio of education programmes

1. The motivation for the Teacher Development Programme (TDP) is based on the recognition that “children in Nigerian schools are not learning” and the quality of teaching, which is of central importance to learning achievement, is a “serious concern” (DFID Nigeria, 2012). Despite large investments in education in recent years, the quality of teachers, teaching, and teacher training continues to be of grave concern for Nigeria (Humphreys and Crawford, 2014). Learning is at the centre of DFID’s approach to aid programming in education, which is based on a ‘learning framework’ for improving learning outcomes principally through “strong and accountable education systems” and “good teachers and great classroom practices” (DFID, 2013).
2. As such, through TDP, DFID Nigeria (DFID-N) will seek to improve the quality of teaching in primary and junior secondary schools (JSS) and in Colleges of Education (CoE), with a focus on six northern states (Jigawa, Zamfara, Katsina, Kano, Kaduna and Niger state). TDP will be implemented in Jigawa, Zamfara and Katsina in the programme’s pilot phase (late 2014) while in the rest in its second phase (late 2016), until the project ends in 2019.
3. TDP’s ‘core business’ is teacher training and there will be several complementarities with other DFID-funded education programmes in Nigeria. Education Sector Support Programme in Nigeria (ESSPIN), which operates in six states including Jigawa (a TDP phase 1 state) and Kano and Kaduna (TDP phase 2 states), has a core focus on school quality improvement, accountability and governance. The Girls’ Education Programme Phase 3 (GEP3), which operates in five states including two TDP phase 1 states namely Zamfara and Katsina, and one TDP phase 2 state (Niger), focusses on creating an enabling environment for girls’ education. Albeit not part of their primary focus, both ESSPIN and GEP3 have programme activities in the area of teacher professional development. For transparent governance and administration of the teaching workforce, as per the programme’s business case, TDP is expected to have complementarities with DFID-funded programmes in Nigeria like SAVI and SPARC, as also with teacher training and development programmes being funded by other development partners such as USAID, British Council and JICA¹ (DFID Nigeria, 2012).

¹ USAID’s Northern Education Initiative (NEI) operates in two Northern states (Sokoto and Bauchi) and provides teacher training within a relatively small number of schools. JICA provides teacher training in Maths and Science in three states (Niger, Kaduna and

1.1.2 State of the Nigerian school teacher

4. There are just over half a million teachers (234,463 women) employed in public primary schools in Nigeria.² These numbers do not include low cost, private, officially unrecognised schools and other private provision in states like Lagos, Enugu and Kwara, nor teachers engaged in Integrated Qur'anic and Tsangaya Education (IQTE) and nomadic education. They also conceal the shortage of women teachers in parts of the north, particularly in rural areas.

5. The shortage of teachers in Nigeria has been widely discussed, and this is particularly acute in relation to shortage of qualified teachers. According to the 2014 Education for All (EFA) Global Monitoring Report (GMR), Nigeria has by far the world's largest gap to fill in terms of additional primary teachers needed. Between 2011 and 2015, it needs 212,000 primary school teachers, 13% of the global total (UNESCO, 2014) which implies, *inter alia*, a commensurate fiscal infusion into the budget for teacher salaries. The report goes on to say that "countries that require additional teachers will have to increase their overall budgets for teacher salaries" and cites the UNESCO Institute for Statistics' new analysis which found that US\$4 billion was required annually in Sub-Saharan Africa to pay the salaries of these additional primary school teachers. This is equivalent to 19% of the region's total education budget in 2011 and Nigeria alone accounts for two-fifths of the gap (UNESCO, 2014). Inevitably, the lack of teachers in many schools results in higher pupil-teacher ratios (PTRs) and overcrowded classes. The recommended PTR at primary level is 35:1, while at junior secondary level it is 40:1 (Federal Ministry of Education, Nigeria, 2004).

6. More detailed data are available from teacher supply and demand studies for Jigawa (Thomas 2011) and Katsina (Bennell 2014, draft) states. Jigawa has a serious shortage of qualified teachers (Thomas, 2011).³ In Jigawa, the PTR for all qualified state primary teachers was 96:1 against a norm of 40:1. Including all teachers (not only qualified teachers, the Jigawa PTR falls to 34:1. The PTR for qualified teachers in JSS grades in Jigawa is 30:1, and 24:1 for all teachers (Thomas 2011: 47). However, these numbers conceal wide disparities in PTR between LGAs within Jigawa. Class sizes are likely to be even larger due to flawed data, to staff deployment practices within schools, or to classroom or staff shortages that may result in two classes being put together (Humphreys and Crawford, 2014). Katsina also has a shortage of teachers at the primary level, where student-teacher ratios on average and in every LGEA exceed the prescribed size of 40:1 (Bennell 2014, draft). In nine LGEAs, student-teacher ratios exceed 100: 1. Moreover, nearly half of primary school teachers were not fully qualified, so student-qualified teacher ratios are likely to be even higher.

7. Aside from shortage of teachers, one reason that is considered the root-cause for the low quality of teaching and learning in Nigerian schools is the lack of qualified teachers. The minimum qualification for teaching at nursery, primary and JSS is the National Certificate of Education (NCE) taught through the Colleges of Education (COE). Education Management Information System (EMIS) data from 2011 shows that almost two-thirds of public primary teachers are qualified, on average, although regional variation is great. The lack of qualified teachers is generally worse in the more rural, northern regions of the country: in the North East and North West well under half the teachers are qualified whereas in the South West almost all teachers are qualified. There is also great variation among states (and LGEAs). For instance, a high percentage (56%) of basic education teachers in Jigawa State are classified as unqualified (64% in pre-primary and primary; 20% in JSS) (Thomas, 2011).

8. However, a 'qualified' teacher, by no means, can be conclusively considered to be an 'effective' teacher. An ESSPIN Teacher Development Needs Assessment (TDNA) showed that out of 19,823 teachers (fully qualified or not) assessed in Kwara, only 75 were able to mark correctly the test papers in Mathematics and English language aimed at 9-10 year olds. In Lagos the numbers were 16 out of 21,358. In Kaduna only 1 out of 1599 and no teacher in Jigawa and Kano achieved the competency level. Qualified

Plateau) with ambitions to cover all Nigerian states. The British Council is also working in the North in support of English language teachers' training.

² According to the Nigerian Digest of Statistics cited in the TDP business case.

³ Qualified is defined in Thomas (2011: 12) as "NCE holders, graduates and holders of other postgraduate teaching qualifications."

teachers performed as badly as the unqualified teachers. In Katsina, the results of the GEP TDNA show that only 2 of the total of 2,281 teachers could be assessed as having achieved a desirable level of professional working knowledge and skill (Johnson and Hsieh, 2014).

9. The basic minimum teaching qualification (i.e. the NCE) consists of three years of training, including a three-month placement in a school, which is in the process of being extended to a one-year internship (NCCE 2005). In-service education is currently primarily aimed at upgrading teachers' professional qualifications (rather than improving teaching), part of a process of professionalisation that is at an early stage in Nigeria. The upgrading of teachers' qualifications to NCE is the responsibility of the National Teachers' Institute (NTI) (Adelabu 2005), while in-service teacher education is coordinated by the Teachers' Registration Council of Nigeria (TCRN), a parastatal of the Nigerian Federal Ministry of Education (FMOE), which is also the teachers' regulatory body, but in collaboration with other agencies and development partners. The upgrading of qualifications occurs primarily through distance learning programmes and vacation and 'sandwich' programmes run by the colleges of education and university institutes of education (Adekola 2007).

10. Low pay and working conditions, as well as low societal status as the cause of low teacher motivation, high absenteeism and attrition have been discussed in the literature (Adekola, 2007a; Adelabu, 2005; Humphreys and Crawford, 2014) as well as spoken about in focus-group discussions (FGD) organised during the preparatory phase of this evaluation framework. (Sherry, 2008), through a teacher motivation and perception study, found that teachers often felt undervalued and neglected by government, having been excluded from the decision-making process and thus having lost trust in government, the main issue being related to pay and work conditions: low pay in comparison to other civil servants and professionals (Adekola, 2007a; Adelabu, 2005; Sherry, H, 2008); further compounded by rural postings, high accommodation and transport costs (Adelabu, 2005; Sherry, H, 2008); delays in salary payments in some states due, to a large extent, to the decentralised and uncoordinated system of government and responsibility (Adelabu, 2005; Sherry, H, 2008); these leading to teachers seeking dual employment until they are paid (Adelabu, 2005); promotions being made on the basis of qualifications and years in service rather than competence further aggravated by lack of agreed teaching standards (Adekola, 2007a; Thomas, 2011); inadequate and unsafe conditions in schools (Adelabu, 2005; Sherry, H, 2008); overcrowded classrooms and lack of teaching materials (Adelabu, 2005; Sherry, H, 2008); and lack of supervision, opportunities for professional development, and support (Thomas, 2011).

11. In-service teacher education is currently being delivered to state primary teachers. It is believed that around 100,000 teachers per year across all of Nigeria are trained for 5 days paid for by federal Millennium Development Goal (MDG) funds (as reported in DFID Nigeria, 2012). Finally about 30,000 teachers are enrolled on NTI's four-year distance learning course to obtain the Nigerian Certificate of Education (although 15% of this number are not actually teaching). In addition, SUBEBs organise needs- and opportunity-based training for practicing teachers in states. These state-led initiatives have often been one-off, limited in scale and resources, and have relied upon different sources including Colleges of Education (CoE), Universities and institutions such as the NTI (DFID Nigeria, 2012).

12. Where teachers undergo in-service teacher education this is usually external workshops rather than in-school support, and there is no system to replace absent teachers (Boulton et al., 2009). Teachers often have to pay out of their own pockets to attend workshops (Adekola, 2007a; Sherry, H, 2008). Teaching practice is too short and inadequately monitored and schools are often selected based on their proximity to save on transport costs rather than on account of the head teacher's ability to provide support to a trainee (Adekola, 2007a; Adelabu, 2005; Humphreys and Crawford, 2014) leading to ineffective in-service training.

13. A number of issues pertinent to the constraints faced by state primary teachers in Nigeria have not been covered in detail in this section. A full review of these issues is beyond the scope of this evaluation framework but readers may find excellent literature elsewhere, such as: (Adekola, 2007a; Adelabu, 2005; Humphreys and Crawford, 2014; Igwe and Rufai, 2012; and Thomas, 2011).

1.1.3 State of Nigeria's colleges of education, teacher-educators and student-teachers

14. Pre-service teacher education takes place in state and federal COEs, universities, polytechnics, and private institutions. The basic minimum teaching qualification (i.e. the NCE) consists of three years of training, including a three-month placement in a school, and is presently being reformed to include a number of changes including an extended one-year placement scheme. The main discussion in this section will centre on the state and federal COEs.

15. The setup of the COEs was widely criticised on the basis of its outdated curriculum and teaching methodology, inadequate facilities and staff shortages, a lack of capacity on the part of some lecturers, the poor calibre of trainees and a lack of monitoring of student-teachers' progress. However, empirical evidence on the teaching and learning processes that ensues in these COEs, including monitoring student-teacher progress in the COE and when student-teachers are on teaching practice, is scarce. See Humphreys and Crawford, 2014 for a full discussion of these issues.

16. The NCE curriculum has come under criticism on several counts. Humphreys and Crawford, 2014 note that the emphasis in teacher education is on the technical (pedagogical) aspect of teacher development, while the equally important social and affective aspects of teaching have merited less attention. The outdated NCE curriculum and teaching methodology, and particularly a theory-heavy pedagogy means that student-teachers graduate with an NCE degree without the essential and practical skills required in school (Adekola, 2007b; Thomas, 2011).

17. Other constraints faced by pre-service training institutions include: overcrowded lecture halls (Adekola, 2007a; Sherry, H, 2008); lack of progress-monitoring and adequate supervision (Allsop, and Howard, 2009); lecturers' own capacity limitations in terms of professional and pedagogical knowledge, whereby a number of lecturers are "graduates without a teaching qualification" (Adelabu, 2005; Thomas, 2011); and finally low calibre of student-teachers as those unable to get admitted in other professional courses enrol for NCE certification and often diverge into non-teaching professions after having acquired the degree (Allsop, and Howard, 2009).

1.2 TDP's objectives and outputs

1.2.1 Programme objectives

18. The Teacher Development Programme (TDP) is a 6-year (2013-2019) project funded by UKAid/DFID to support the Federal and State institutions responsible for the pre- and in-service training and development of basic education teachers in six states in Nigeria. The overall goal of the TDP is to improve student learning in the target schools through changing the way children are taught in the classrooms under basic education (P1-6 and JSS 1-3). The Programme will work in up to six states (beginning with Jigawa, Katsina and Zamfara in late 2014 and later to three additional states – Kano, Kaduna and Niger State).

19. The main objective of TDP is to improve the skills of teachers in the three core curriculum subjects of English, maths, and science & technology and produce better teachers through a combination of pre-service and in-service interventions. The specific objectives of TDP are to: (a) improve in-service training of primary and junior secondary school teachers; (b) develop more effective teacher educators of primary and junior secondary school student-teachers through system reform; and (c) strengthen evidence-based research to influence and inform decisions on teachers' effectiveness and efficiency (TDP, 2014a). The division of total resources to the components of TDP are ca. 80% in-service, 10% pre-service, and 10% results and evidence. There are three outputs described further below.

1.2.2 Programme outputs

1.2.2.1 Output 1: In-service teacher training

20. TDP's in-service training component aims to help the participating states to:
 - i. establish a school-based, cost-effective in-service training programme supported by the 'trainer in-the-pocket' model (described later) and similar approaches; and
 - ii. establish permanent cadres of teacher educators responsible for the coordination, development and delivery of in-service programmes.
21. Through these actions TDP targets States to provide regular and on-going training and support to a total of 62,000 teachers by 2019. (TDP, 2014b)
22. TDP's in-service teacher training model is grounded in three key principles, referred to as 'three pillars' in the Programme's in-service strategy document, that drive the design of the intervention and activities within the area of in-service teacher training (TDP, 2014b). These three pillars are:
 - school-based interventions: collectively termed as 'New Classroom Activities' which comprise teacher training, materials for students and teachers, and teacher support;
 - 'trainer in the pocket': Access to audio-visual resources, anytime and anywhere, ensured through the use of mobile technology; and
 - continuous support: ongoing support to teachers over a few years through multiple layers and mechanisms (workshops, cluster meetings, classroom observation, peer support, self-study materials etc.) instead of once-off training. The support mechanism for continuous professional development is to be institutionalised within the school and state.
23. These in-service activities will roll out in phase 1 states by selecting and training two teachers for English and two teachers for maths on p1-3 materials in year 1 (2014), and then in year 2 (2015) training these same four teachers on p4-6 materials, and training two of these teachers in p4-6 science&technology materials. Further details of in-service activities are discussed in Annex B (B.1).
24. The overall model for the TDP in-service component is to establish an effective model for in-service teacher training (the three pillars of 'New Classroom Activities', 'trainer in the pocket', and continuous support referred to above), and then to work with other DFID education programmes (ESSPIN and GEP3) and with national and state government education institutions to apply this model in TDP states and beyond (though these latter objectives are not captured explicitly in the TDP theory of change).
25. Two aspects of the in-service design generate difficulties for the evaluation framework at this point. First, materials for the 'New classroom activities' for students and teachers, and for the trainer in the pocket are still being developed, which means that precise questions around these materials cannot be developed at this stage. Second, anecdotal evidence suggests that the allocation of teachers to classes varies in TDP primary schools: some teachers teach a single subject (e.g. only English), some a single grade (e.g. only Grade 1), and some a single cohort. This means that the transmission mechanism from improved teacher training to improved learning will vary across schools. Our proposed solution to this problem – constrained randomisation – is outlined in section 4.1 below.

1.2.2.2 Output 2: Pre-service teacher training

26. The pre-service training component is not a focus of this evaluation other than where it has synergies with the in-service component. We briefly outline the pre-service component in this section, highlighting these expected synergies.

27. The pre-service component of the TDP has six main objectives. It will (a) support the implementation of the new NCE curriculum⁴; (b) facilitate the restructuring of COEs consistent with the new NCE curriculum; (c) support the enhancement of quality assurance processes and procedures in the COEs; (d) support National Commission for Colleges of Education (NCCE) quality assurance processes and procedures; (e) enhance active learning pedagogies and update subject knowledge of COE staff; and (f) enhance quality of teaching practice in the COEs. (TDP, 2014c)

28. A number of activities will be undertaken in order to implement the pre-service component of TDP. These will include:

- Training of master-trainers on the use of low-cost mobile technology (akin to the “trainer-in-your-pocket” technology to be used in the in-service component of TDP), with possible synergies with the in-service component in supporting further in-school training;
- Supporting COEs to conduct the training for 4000 student-teachers to use this low-cost mobile technology, with synergies as above;
- Supporting the development of a mentors' manual and organising the training for mentors and supervisors (teacher-educators) on the revamped Teaching Practice model;
- Supporting COEs to implement the new approach to Teaching Practice with strong supervision and mentoring;
- Supporting COEs to create “Associate Schools” that will engender closer links between schools and COEs for the mutual benefit of both, with possible synergies with in-service training if those schools overlap with schools receiving in-service support (this is to be decided by the TDP);
- Supporting COEs to train teacher-educators and student-teachers on the Teaching Practice toolkit developed jointly by NCCE and Open University UK; and
- Providing incentives for colleges to improve the quality and effectiveness of teaching practice. (TDP, 2014c)

29. The pre-service implementation strategy notes several synergies between the pre- and in-service components of TDP. “Trainer in pocket”, the model and its associated technology and materials, will be adapted and used as part of the capacity building programme of the pre-service component. Similarly, as part of its attempt to improve the effectiveness of teaching practice, the pre-service component will support colleges and SUBEBs to establish “Associate schools” in all the three states. It is planned that such schools will mainly, if not exclusively, be those the in-service component is already working with. This will facilitate the sharing of materials but more importantly gives the two components the opportunity to draw from each other’s strengths and successes, learn from their respective challenges and try out new ideas and practices in more or less the same setting and thereby reinforce their respective activities.

30. Further details of pre-service activities are discussed in Annex B (B.2).

1.2.2.3 Output 3: Results and evidence

31. While the evaluation framework will not apply to the results and evidence component of the TDP, the evaluation will use some materials generated by this component to support the evaluation of the in-service component of the TDP.

32. There are three elements of the TDP Results and Evidence (R&E) strategy, namely:

- The evaluation of the Programme in terms of its impact, outcomes and outputs;

⁴ Presently limited amounts are known about the contours of the new NCE curriculum. The new NCE curriculum requires colleges to deliver programmes in a selection of up to five areas of specialisation (ECCDE, Primary Education, Junior Secondary Education, Special Education, and Adult and Non-formal Education) as a consequence of which colleges will need to restructure. NCCE has issued guidelines on restructuring which envisage the formation of Schools of ECCE, Primary, etc with staff being allocated to each school according to the specialism they will teach. Thus, for instance, a member of staff will be located within the School of Primary Education rather than, as at present, within, say, the School of Social Studies.

- Studies that evaluate the results of the Programme but which are intended to add to the educational evidence base in Nigeria more broadly; and
- The monitoring of Programme activities to ensure the quality of their implementation. (TDP, 2014d)

33. The TDP R&E strategy document (TDP, 2014d) details TDP's plans for conducting impact evaluation, perception studies; and setting up within-programme monitoring systems—drawing on both R&E activities of existing DFID-N education programmes (e.g. from ESSPIN and GEP3) as well as generating new data and evidence.

34. Further details of activities related to the R&E strategy are discussed in Annex B (B.3).

1.3 Evaluation purpose

35. Consultations with DFID and the TDP in Nigeria indicate that the evaluation should focus on the in-service component of the TDP, and should not evaluate the pre-service or results and evidence components of the TDP, except where there are clear synergies with the in-service component. The evaluation should also focus on the TDP phase one in Jigawa, Katsina and Zamfara (2014-2016) rather than phase two in Kaduna, Kano and Niger (with detailed planning from January 2016 and implementation from September 2016 to 2019).

36. There are three potential purposes for the TDP evaluation: i) formative (contributing to the implementation of TDP), ii) summative in terms of ensuring accountability for the TDP's performance (does the TDP deliver what it is supposed to sustainably, and is this the right model for improving teacher performance and learning outcomes where the TDP is working?), and iii) learning from TDP for what works in improving teacher performance in other contexts (in Nigeria and elsewhere). These are taken in turn, and in order of priority.

1.3.1 Learning from the TDP for other contexts

37. The priority for the evaluation is learning about the in-service component. This priority results from the lack of clear evidence on what can be done to improve teachers' knowledge and practice and subsequently pupils' learning. For example, a recent DFID systematic review of evidence on teachers was not able to draw on large amounts of evidence on the causal relationship between in-service training and teacher competence, particularly in resource-poor environments (Westbrook et al 2013).

38. DFID Nigeria's priority for the TDP evaluation is therefore around learning what in-service activities work to support teachers in resource-poor environments. This implies a rigorous assessment of changes in teacher competence and learning outcomes that can be attributed to the TDP, and an explanation of why these changes take place (or not). This suggests an experimental or quasi-experimental evaluation design (to generate attributable estimates of impact in each state) as part of a theory-based evaluation (to explain what works and why). This would be coupled with a process evaluation of whether and how the TDP in-service programme is achieving its outputs. This learning should be used to influence partners in other States in Nigeria, but also in other resource poor environments. Estimates per state are important as each State is very different, and may receive slightly different programmes depending on interactions with other DFID education programmes.

39. The implication of this requirement for rigorous data on quantitative changes in pupil learning and teaching caused by the TDP's in-service output is that a major part of the evaluation will involve quantitative surveys on teaching and learning. The resource implications of these surveys are discussed below in section 3.2. If the main purpose of the evaluation were accountability, the balance of evaluation resources would reflect the funding: 8:1 in-service:pre-service. However, the main purpose of the evaluation is learning, so resources will be devoted to the areas where learning will be most useful: the in-service component.

40. The learning aspects of the evaluation should also consider synergies between the in-service and pre-service components. These include the use of in-service materials and technology for pre-service training, the creation of associate schools that would contain teachers from both programmes, the links between the components created by the teacher facilitators, and the extent to which teachers need to have taken part in good pre-service training to benefit fully from in-service training. The evaluation should consider whether these synergies have been productive for improving teacher competence.

1.3.1 Formative evaluation to help the TDP learn

41. Both DFID and the TDP also see significant formative value in this evaluation. Specifically, the evaluation should test in the TDP phase one whether the in-service model for improving teacher competence works overall, which parts work well, which parts work less well, and why. The purpose of this part of the evaluation is to inform phase two. This means that the evaluation should test changes in teacher competence that result from in-service work, and provide initial explanations for why changes are occurring, with evidence available by early 2016 (meaning data collection in 2015, or the TDP year 2).

42. This formative evaluative work is not expected to include tests of student learning, because these outcomes typically take longer to materialise, whereas improvements in teacher competence should be evident more swiftly. This evidence would be formative: it would influence the way in which the TDP phase two is designed and implemented, with a view to maximising the opportunities for learning from the evaluation.

43. At present, the expectation of formative evidence on whether the TDP improves teacher competence after one year is not reflected in the logframe, which contains milestone targets for outputs in year 2 (e.g. number of teachers trained), but not outcome targets (e.g. % change in teacher pedagogical competency, which is expected to increase by 5% by year 3 (2016)). If the importance of formative information teacher competency is accepted, this would ideally be reflected in the logframe.

44. The TDP state teams (and, to some extent, DFID-N) see formative value in ensuring that the TDP programme participants (teaching staff, parents and students) are involved in the evaluation. This is because their active support and participation is critical to the success of the programme, and therefore eliciting, understanding and responding to their views should inform the evolution of the programme. However, the engagement of these groups is beyond the scope of the evaluation team's resources. It is proposed that the TDP engage these individuals so that formative value of this evaluation is maximised (i.e. the likelihood that the evaluation contributes to the TDP's outcomes is improved).

1.3.2 Evaluation for accountability of TDP's performance

45. Neither DFID nor the TDP see significant value in an evaluation for accountability. Rather, this will be provided by the annual reviews of the programme, which are also managed by EDOREN.

2 The TDP's theory of change

46. This section unpacks the TDP's theory of change, in order to inform the evaluation questions. The theory of change of a programme articulates a logical chain from the programme's activities to its eventual goals. The theory of change is a very useful tool for an evaluation that seeks to set out whether a programme is achieving its goals, and if not, why, because the evaluation can test each of the links in the logical chain. Some of the steps of this chain form the logical framework and associated assumptions for a programme, but others (usually less easy to observe) are not included in the logical framework.

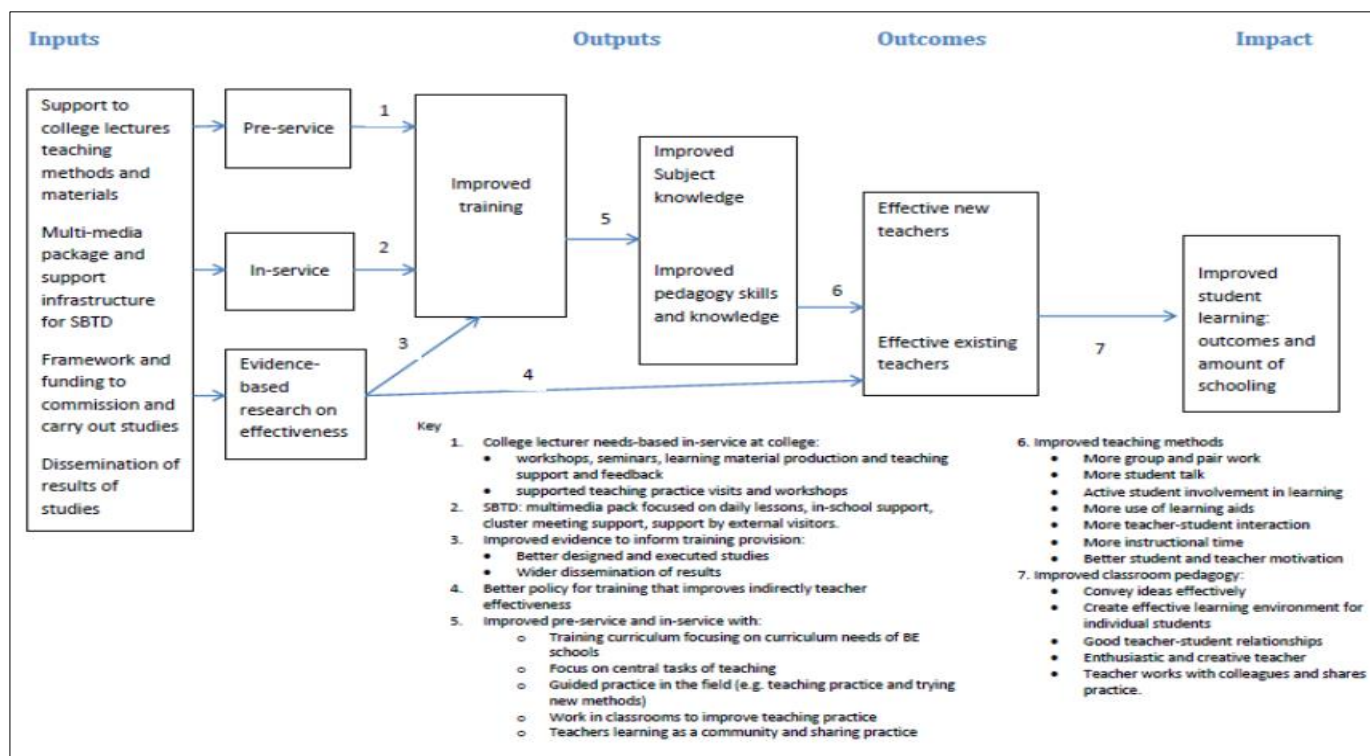
47. The TDP has both a theory of change and a logical framework for internal and accountability purposes. This section starts from the TDP's theory of change as set out in the TDP documentation, and then extends it to clarify some additional links in the chain for the in-service component that may be important but are not included in the documentation. Figure 1 illustrates the TDP's theory of change as articulated in the TDP business case and we briefly summarise the assumptions and evidence underlying various links between inputs, outputs, outcomes and impact for each of its components (in-service, pre-service, and research & evidence) in this section. This relates to the TDP's logical framework provided as Annex C. In the next section, we provide more detail on the theory of change for the in-service component.

48. **TDP output 1 - In-service training:** the theory underlying the in-service training component is that students learn more when they are taught by competent and effective teachers, and that teachers become more skilled and knowledgeable (both in terms of pedagogy and subject-knowledge) through training. Training effectiveness is improved through head teacher support, mentoring and supervision, peer interaction and support from similarly trained teachers where teachers are able to support each other, and regular involvement by visiting trainers/mentors/supervisors. The most effective training focuses on classroom behaviour for both the teacher and the students, supported by learning materials for use with students, and standards of teaching techniques for teachers. Other things being equal, pupils' satisfaction and interest in schooling improves as teaching quality is enhanced and this positively affects their attendance, achievement and reduced drop-out rates. As a combined result of all this, 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.

49. **TDP output 2 - Pre-service training:** the theory underlying the pre-service training component is that teachers leaving COEs and entering the teaching profession with greater subject knowledge and pedagogical competence (further developed later through in-service training) become better teachers than those teachers with inadequate pre-service training. Further, educating student-teachers is more effective where COE lecturers focus on classroom practice in their college-based work with trainees; where these student-teachers undertake well-designed teaching practice components in schools; where the curriculum (NCE in this case) is attuned with actual pedagogy and subject-based demands on teachers in schools; and where there are thorough quality assurance processes and procedures for certification of new teachers. Improved teaching practice involves preparation and support to trainees when they are in schools during their COE placement.

50. **TDP output 3 - Research and evidence:** the theory underlying the R&E component is that a greater understanding of the factors influencing teachers' performance, constraints leading to low teaching quality, etc. can be informed by empirical studies and research commissioned by this programme. This will inform TDP, DFID-N, and relevant stakeholders whether the intended logical chain from inputs to outputs to outcomes and impacts is being materialised as described in section 2 below. These will not just test the internal validity of TDP's theory of change but also provide valuable evidence for consideration in other teacher education programmes in Nigeria and elsewhere. This evidence will be crucial for further iterations of the TDP, as also potentially in decisions around teacher recruitment, deployment, assessment and career development. The findings of this analysis will enable decision makers to reinforce and sustain the intended impact of this programme by ensuring a more optimal use of the scarce, good quality teaching resources.

Figure 1 TDP's theory of change



Source: TDP business case, page 22.

2.1 Unpacking the TDP's in-service theory of change

51. As indicated above, DFID have asked that the evaluation framework focus on the in-service component of the TDP, together with any synergies with the pre-service component. This section therefore unpacks the in-service component's theory of change in more detail, setting out more fully the causal links between TDP in-service activities and the expected impacts of improved student learning outcomes and amount of schooling. We first describe the TDP's planned in-service strategies in more detail, and then organise these strategies around a theory of change.

2.1.1 TDP in-service strategies

52. The in-service component of the TDP envisages five core strategic areas, as per the TDP in-service teacher training strategy:

- Collaboration and partnership
- Training and support
- Materials development
- Technology use and management
- INSET programme implementation

53. Under **collaboration and partnership**, the TDP aims to i) take the lead in developing a model for standardised teacher training, and then coordinating the application of this model across the TDP, ESSPIN and GEP3, ii) develop strategic partnerships with the FMOE, UBEC and NERDC to ensure the national roll out and long-term sustainability of the in-service training model, and iii) partner with SUBEBS that will become key strategic homes for the TDP (because reforms to CoEs that would see them lead in-service teacher training have not come to fruition).

54. Under **training and support** the TDP aims to i) build a whole-school development approach by gradually training and supporting teachers in English, Maths, and Science and Technology, ii) 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, iii) engage head teachers actively in pedagogy and leadership and management training, iv) recruit a pool of teacher educators as the teacher development team, v) recruit teacher facilitators from the current school supervisor cadre, and vi) design and implement separate professional development programmes and activities for teachers, head teachers, teacher facilitators and teacher development teams.
55. Under **materials development**, the TDP aims to i) develop new or adapt existing materials for teachers, teacher facilitators and head teachers, ii) develop a common pedagogical framework with content that cuts across all subject areas and work strands, iii) ensure the relevance of TDP materials and content to text books used in each state while aiming to ensure there is as much common content as possible, iv) use ESSPIN lessons plans for the basis of TDP materials for English and Mathematics for P1-5, v) adapt head teacher leadership and management materials from ESSPIN and GEP3, vi) engage with the Education Resource Development Council at state level for teacher training material development and strengthen its capacity to become the custodian of all audio-visual resources within and beyond the project period, vii) engage local experts in developing materials, viii) implement materials development based on a step-by-step approach, ix) produce Hausa based instructional materials for teachers on maths P1-3, x) use English as the language of instruction for all materials from P4-6 and JSS1-3.
56. Under **technology use and management**, the TDP aims to i) select appropriate technology based on an assessment against a range of pre-defined criteria, ii) produce audio-visual resources on secure digital (SD) cards, iii) ensure state contributions for procurement and management of technology devices, and iv) schools are custodians of technology devices but teachers can take and use the devices outside school hours.
57. Under **INSET programme implementation**, the TDP aims to i) implement a two year pilot with 6,000 teachers from 1,500 schools across all LGEAs in three initial states before scaling up to six states, ii) ensure LGEAs, clusters, schools, teachers, teacher facilitators and Teacher Development Team members are selected and recruited based on clear agreed criteria, iii) ensure the participation of the SUBEB and other relevant authorities in the selection process, and iv) support states to expand teacher support to other schools in the state from 2015 onwards.
58. These activities are intended together to lead to the TDP's objectives of improved learning outcomes within TDP schools and beyond. The next sections describe the causal chain between strategies and these objectives in more detail.

2.1.2 TDP strategies within a theory of change

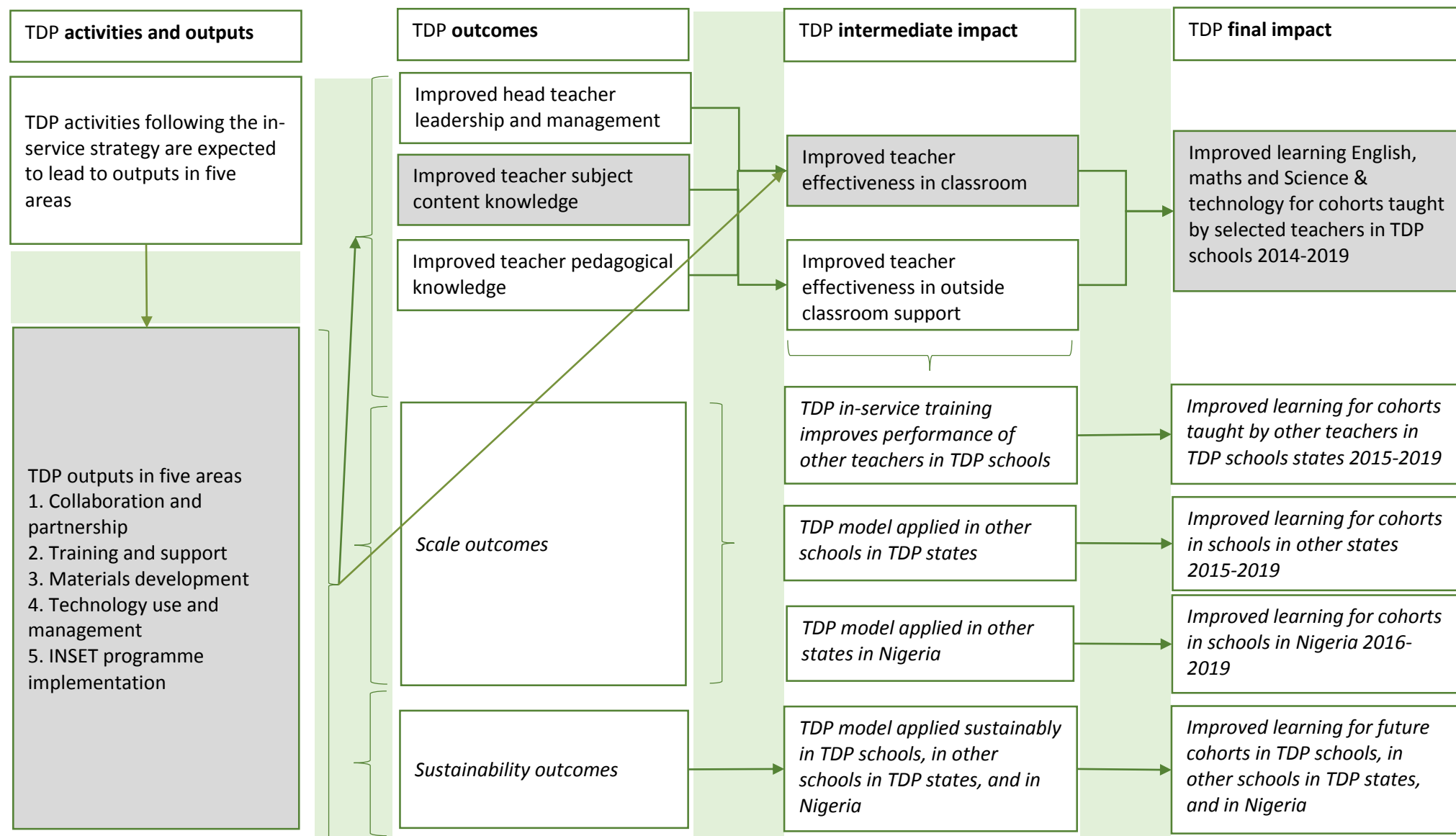
59. The text below provides more detail on the causal mechanisms linking the TDP's in-service activities to its expected impact, and the assumptions around them. The text is intended to shed light on these mechanisms in order to more precisely evaluate the programme; it is not intended as a criticism of the design (though it may be of use to the programme in planning around risks). Rather, the TDP appears to offer an innovative solution to an entrenched problem in a very challenging context, and it is our hope that the theory of change analysis offered here will support this solution and allow us to assess whether it has worked.
60. We divide the TDP theory of change into three parts. The first concerns the causal chain between TDP outputs and the primary objectives of improving learning outcomes in TDP schools. In terms of the OECD DAC evaluation criteria (see section C.1), this part forms the basis for an evaluation of impact, effectiveness and relevance. The second concerns the causal chain between TDP activities and the secondary objectives of achieving improved learning outcomes in other schools in TDP states and in Nigeria. This second part is not the subject of explicit targets in the logical framework, but is an objective

that has influenced the design of the in-service strategy and against which the programme can be evaluated, particularly in terms of sustainability. The third part looks in more detail at how TDP activities are expected to lead to outputs in terms of improving learning outcomes, scale and sustainability. This last part would form the basis for a process evaluation to assess efficiency, and will not be explored in detail here.

61. Figure 2 below presents the three more detailed aspects of the TDP in-service theory of change with: activities to outputs on the left, outputs to final impact in the top half, and scale and sustainability in the lower half (in italics). The activities to output part of the theory of change would be the subject of a process evaluation. The outputs to impact part of the TDP theory of change is that i) the TDP outputs in the five areas listed above will lead to improved teacher and head teacher skills and knowledge (TDP outcome), ii) improved teacher skills and knowledge will improve teacher effectiveness inside and outside the classroom (TDP intermediate impact), and iii) improved teacher effectiveness will improve learning outcomes for students taught by those teachers (TDP final impact). In addition, materials will help teachers' effectiveness in class (depicted by the long arrow from outputs to teacher effectiveness in the figure). The scale and sustainability part of the theory of change reflects TDP's ambitions to have an impact in other places and over a longer period of time than the TDP project. Each link in this chain has a number of assumptions and other factors that influence the transition from one stage to another. The following sub-sections break down this causal chain further as a basis for formulating evaluation questions and approaches to answering them, looking at:

- Intermediate impact to final impact;
- Outcomes to intermediate impact;
- Activities and outputs to outcomes;
- Scale and sustainability outputs to final impact.

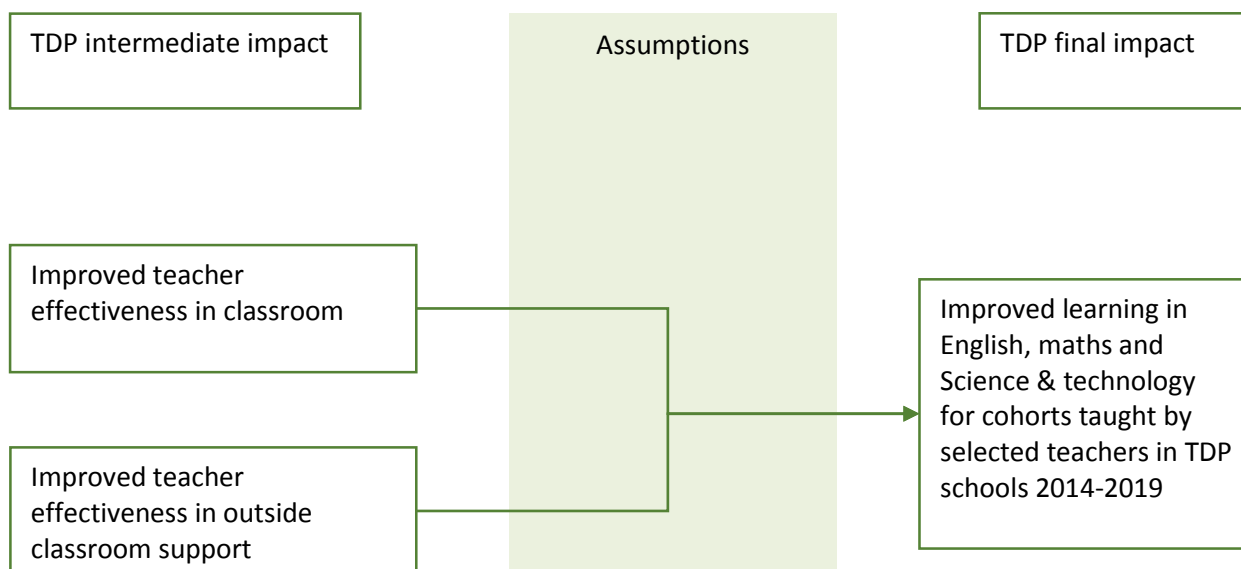
62. The diagrams below are not intended to be equivalent to the TDP logical framework. The diagrams below are intended to represent a theory of change – what we would expect to happen as a result of TDP activities. The TDP is not accountable for and has not committed to achieving these results. The logframe results – to which the TDP has committed – are shaded in the figure below.

Figure 2: TDP theory of change from activities to impact, scale and sustainability⁵⁵ Boxes relevant to the logframe have been shaded.

2.1.2.1 Intermediate impact to final impact

63. Drawing on DFID's learning framework (DFID 2013) and literature on education production functions (Hanushek and Rivkin, 2006), the transition from intermediate impact (**improved teacher effectiveness**) to final impact (**improved learning outcomes**) involves several assumptions, shaded in the figure below.

Figure 3: TDP intermediate impact to final impact



64. The transition depends on the cumulative impact of the following assumptions:

- Children attending school regularly;
- Children having the capacity to learn from improved teaching in the language of instruction (they are school ready);
- Children receiving adequate support for learning at home;
- Children supporting their peers to learn;
- Teachers attending class regularly;
- A class size small enough to allow improved teacher effectiveness to have an impact;
- Adequate classroom materials (blackboards, books, desks, etc.) being available; and
- Curriculum and materials that are appropriate to the language and ability of students.

65. Overall, there is good evidence that 'effective' teachers are significant determinants of learning achievement among pupils (Aslam and Kingdon, 2011; Hanushek and Rivkin, 2006; Nannyonjo, 2007; Westbrook, 2013). Furthermore, this effect is both cumulative and residual such that a better teacher in one particular year results in better test scores in that year and, in fact adult-life outcomes, in later years (Chetty et al., 2011). Finally there is strong evidence that the impact of other investments in the education sector are dependent on the quality of teachers (Hanushek and Rivkin, 2006; Yoshikawa et al., 2007). This suggests that in general the link between effective teaching and enhanced learning outcomes is strong, so that if the TDP's in-service training is able to improve teacher effectiveness, learning outcomes should improve, other things being equal.

66. However, there are four reasons why the link between teacher effectiveness and improved learning is problematic for the TDP and its evaluation.

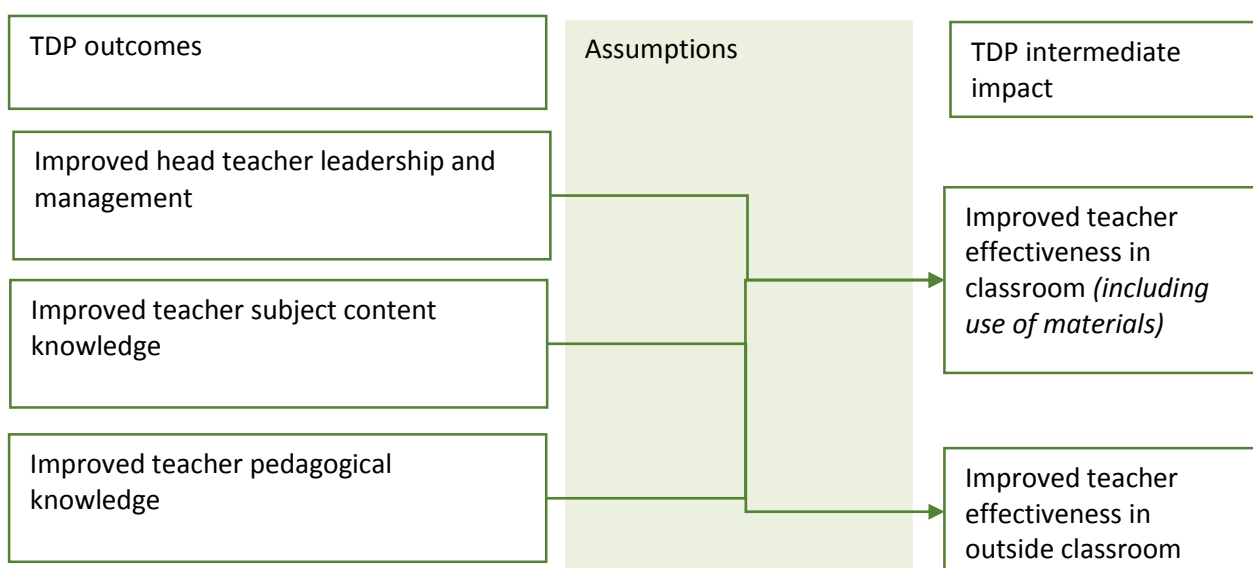
- i. The many other factors listed above may limit improvements to learning in Nigeria, because the evidence suggests that few of these conditions are met. This implies that improving teacher effectiveness may not be sufficient (though it may be necessary) to improve learning outcomes. A recent review of the literature on basic education in Nigeria (Humphreys and Crawford 2014) indicated that there is a wide range of constraints to improved learning outcomes, including teacher effectiveness (meaning competence and time-on-task) but also a range of different inputs (books, materials, etc.), student-teacher ratios (see also teacher supply and demand studies by Thomas 2013 and Bennell 2014 discussed above), the curriculum and language of instruction, and various demand-side factors such as school attendance rates, household characteristics, and conflict. The low school attendance rates would be a significant constraint to improving learning outcomes of the entire school-aged population; given that many children do not attend government schools regularly, the TDP approach is very unlikely to affect them. Of the other factors, it is not a priori obvious which of these factors has the largest influence on education outcomes in Nigeria, and there is no evidence available to make this case clearly.
 - iv. Given this, there may be challenges in identifying the contribution of improved teacher effectiveness in an evaluation. The evaluation needs to control for these other factors, ideally through an experimental evaluation design, so that the impact of TDP's improvement in teacher effectiveness (assuming this occurs) on learning outcomes can be isolated. Even after capturing various observed family, community, teacher, and other inputs, we are still left with a situation whereby we are heavily relying on a small set of observable characteristics to define teacher quality, usually on teacher qualification and experience, even when studies repeatedly show that these observed characteristics explain very little actual variation in teacher effectiveness. As such, given data limitations, unobserved attributes determining teacher effectiveness cannot be captured in these models easily even though they may dictate the effect TDP may have on teacher effectiveness.
 - v. Improved teacher effectiveness is most likely to improve learning outcomes in a demonstrable way if students are repeatedly taught by TDP-trained teachers. In other words, if a student is taught by a 'more effective' teacher for only one year, their learning will not improve as much as it would if they were taught for their entire school career. In the TDP states, some teachers teach a grade (e.g. P1), some a subject (e.g. English) and some a cohort. The impact on learning would probably be largest for the latter set up, but this will vary, making the identification of a single % impact on learning outcomes problematic.
 - vi. Finally, it is possible that improving teacher effectiveness may have other impacts that limit improvements in learning outcomes (such as good teachers moving schools or many new students attending schools with good teachers thereby overusing scarce resources). This is more difficult to deal with. There are two main possibilities. First, that teachers that improve leave their schools because they are able to obtain better jobs, and the benefits are not therefore retained by their pupils. This is likely to be a risk particularly in more remote schools, which find it harder to retain teachers in Nigeria. Second, that TDP school-based interventions make TDP schools more attractive for students currently not attending, or currently attending other schools, and the increase in attendance rates overstretchers already scarce resources and reduces overall learning outcomes. There is evidence that improved teacher effectiveness will improve student attendance (Orr et al 2013).
67. This is both a research 'problem' of statistical identification for causal attribution as much as a policy-design conundrum. The most obvious complication to research arises from the fact that observed learning outcomes represent the effect of several interrelated choices – those of parents, teachers, administrators, community, etc. This complexity makes it difficult to separate the various influences reliably. Thus, for example, judging variations in teacher quality require distinguishing teacher effects from elements of students and parents themselves, which in turn requires various econometric designs using panels, instrumental variables, natural experiment or a carefully designed randomised trial with an adequate sample size.

68. The policy design issue arises from the limitations in the number of domains in which we can realistically intervene at a given point in time, and how the impact of a well-designed programme which focusses on one of the domains in the education production function above, say teacher quality in this case, could still be eroded by lack or poor quality inputs in other domains. Ultimately then, it is the complex interplay of these various domains (both observed and unobserved) which determine what the learning outcome for a pupil in a certain grade would look like, making any ex-ante assessment of TDP's likely achievement stand on weak grounds.

2.1.2.2 Outcomes to intermediate impact

69. We next consider the assumptions underpinning the link between **TDP outcomes** (improved head teacher and teacher knowledge) and the **TDP intermediate impact** (improved teacher effectiveness), as shaded in the figure below. This chain is also supported by materials that will be used directly in the classroom.

Figure 4: TDP outcomes to intermediate impact



70. This link depends on:

- Materials being appropriate and available (considered further under outputs below),
- Selected teachers being sufficiently intrinsically and extrinsically motivated to apply their new knowledge,
- Selected teachers being retained in schools where the TDP is operating,
- Selected teachers being class ready, in other words have the capacity to apply their new knowledge,
- Selected teachers being supported to apply their new knowledge, and
- Head teachers being motivated to lead and manage teachers well.

71. The overall assumption around is that improved teacher knowledge will lead to improved teacher effectiveness even if factors that the TDP does not try to control remain unchanged. There are three difficulties with this broad assumption:

- Some of the influencing conditions above are probably not met in northern Nigeria and not all are addressed by TDP activities. For instance, teacher turnover is high, teacher morale is low, many teachers are not class ready (though the TDP pre-service activities may improve this), and management practices are often problematic (see Bennell 2014 and Humphreys and Crawford 2014). While the TDP in-service activities may address some of these, they will not address all. For

example, the TDP's work to train head teachers on leadership and management may improve head teacher motivation and their day-to-day support of teachers to apply their knowledge, but this will not necessarily directly affect teachers' intrinsic or extrinsic motivation, retention, or class readiness. These latter factors will be influenced in turn by overall systems of teacher recruitment, training, deployment, and remuneration. The TDP recognises the importance of this issue, and is convening meetings with government institutions to discuss it – though it remains outside the TDP's direct control.⁶

- ii. If these conditions are not met, it is not a priori clear that improved knowledge will lead to improved effectiveness and measuring any improvement would need to control for these other factors where possible (though many of them, such as teacher intrinsic motivation, are not easily observable). In other words, these other influencing factors may be 'binding' in terms of teacher effectiveness: teacher effectiveness may not be able to improve substantially unless these other conditions also reach a certain level. The international evidence is not entirely clear on this, but a recent thorough review of the evidence related to in-service teacher training on pedagogy (Westbrook, 2013), summarised: "Findings on the effectiveness of continuous professional development [CPD] , however, remain inconclusive: while there is a shift towards teachers' increased awareness and understanding of student-centred pedagogies or use of ICT, these practices are not always immediately implemented in classrooms."
- vii. Finally, these conditions may change during the TDP project implementation process, which may influence outcomes that the TDP is trying to change. For example, exogenous (i.e. not related to TDP) changes to teacher training, remuneration or recruitment could have large influences on teacher motivation and effectiveness.

2.1.2.3 TDP outputs to outcomes

72. The link between TDP outputs and TDP outcomes is complicated, as a range of different outputs in strategic areas co-contribute to different outcomes. Some of the detail of the outputs is still being developed by the TDP, so this section will need to be enhanced during the evaluation. This section looks at the links between outputs and the outcomes in the upper half of Figure 2 (again, note that these are not identical to the logframe outputs):

- Improved head teacher leadership and management;
- Improved teacher subject content knowledge; and
- Improved teacher pedagogical knowledge.

73. The principal outputs contributing to improvements in these outcomes are under in-service strategic areas of **training and support** and **INSET programme implementation**. The other strategic areas are either in support of these (**materials development, technology use and management**) or more directly concerned with scale and sustainability (**collaboration and partnership**). In addition, materials and technology will be directly applied in classrooms.

74. In brief, these improvements are expected to come about through i) the establishment of a system of ongoing training and support in subject knowledge, pedagogical competence, and ii) the development of a model and materials for this training and for teachers to apply in class. In the first two years of the TDP, this will be applied to teachers of primary grades 1-6.

75. The details of the **training** modality are still being developed, but according to the INSET strategy of January 2014, training will be implemented in a cascade model with continuous support and peer-to-peer

⁶ An EDOREN thematic study will produce findings on the influence of teacher management systems on teacher effectiveness in two states by March 2015.

learning. This will be based on a 'cluster' of schools arranged in a geographical location to create a peer network, and will link to ESSPIN's work on head teacher training.

76. The highest level of trainers is the Teacher Development Team, which the TDP will help to set up on a permanent basis to provide continuing support. This group of teacher educators will be managed by the SUBEB and will need either to be developed from the State Support Improvement Teams (SSIT) in ESSPIN states or as a parallel group. The Teacher Development Team would receive training and support (it is currently not clear how this will work), and would then train the next level of teacher educators (the Teacher Facilitators Team) and occasionally teachers and head teachers themselves. It is expected that the Teacher Development Team would provide:

- six days' initial training for the Teacher Facilitators Team,
- bi-monthly meetings with Teacher Facilitators,
- a minimum of eight school visits per month for classroom observations, and
- reporting back to the TDP monitoring and evaluation team.

77. The second level of the cascade model is the Teacher Facilitators Team, based at LGEA level and drawn from an existing cadre of school supervisors/school support officers/quality assurance officers. The Teacher Facilitators Team would support teachers and head teachers directly in schools and classrooms, providing:

- 2.5 days' initial subject-based training for teachers,
- Monthly cluster meetings with teachers on each subject,
- Visits to 4-5 schools and 20-25 teachers per month in classrooms, and provide support before and after lessons,
- Reporting back to the TDP monitoring and evaluation team.

78. These training sessions will draw on various kinds of materials and technological support provided by the TDP. Specifically, the TDP will develop and produce materials on:

- Print and audio-visual materials in English for students for English grades P1-6 and (from September 2016) JSS1-3;
- Print and audio-visual materials in Hausa for students for Maths grades P1-3 and in English for grades P4-6 and (from September 2016) JSS1-3;
- Print and audio-visual materials in English for students for Science&Technology grades P4-6 and (from September 2016) JSS1-3;
- Printed lesson plans and audio-visual teacher guides linked to lesson plans and textbooks for teachers in English for English grades P1-6 and (from September 2016) JSS1-3;
- Printed lesson plans and audio-visual teacher guides linked to lesson plans and textbooks for teachers in English for Science&Technology grades P4-6 and (from September 2016) JSS1-3;
- Printed lesson plans and audio-visual teacher guides linked to lesson plans and textbooks for teachers in English and Hausa for Maths grades P1-3, and in English only for grades 4-6 and (from September 2016) JSS1-3;
- Leadership and management for head teachers (adapted from ESSPIN and GEP3),
- Printed and audio-visual materials for the Teacher Facilitator Teams on pedagogy, teacher training methods, and supportive supervision and mentoring.

79. Each head teacher in TDP schools attending the pedagogy training for one subject, participating in leadership and management training, and participating in head teacher professional development programmes. The specific expectations are that head teachers will i) provide the necessary support to the

trained teachers to apply changed pedagogy in the classrooms including the use of materials, ii) monitor progress of individual teachers, iii) organise periodic staff development meetings. Including the head teacher, four teachers will be selected from each school (by head teachers and LGEA staff) to attend training.

80. The audio-visual materials will be produced on secure digital cards to allow wider distribution and usage. Teachers and head teachers participating in training will be provided with a 'technology pack' (probably a phone or a tablet) that will be held by schools but should be made available for teachers to take and study at home. This is the innovative 'trainer in the pocket' approach that has been used in English in Action in Bangladesh.

81. The link between the training modality and the materials development (TDP outputs) and improved teacher and head teacher knowledge and skills (TDP outcomes), and teacher effectiveness (intermediate impact), depends on several assumptions. Chief among these are that:

- Teacher educators and teacher facilitators are able to absorb and then transfer the skills offered in the TDP training;
- Teacher educators and facilitators have sufficient time and material support to provide support to schools (i.e. they are not directed to other tasks, transferred, or made redundant);
- Head teachers have appropriate incentives/motivation to apply their new knowledge in support of teachers;
- Teachers and head teachers attend training regularly (i.e. they have the time and are managed and supported to so, and not transferred during the training process);
- Teachers can access and use the audio-visual materials (i.e. the technology works, can be charged, is not lost, stolen or broken, is upgraded or fixed where appropriate, can be understood, etc.);
- 2.5 days' training, monthly cluster meetings, and the support visits and mentoring are sufficient to instil new pedagogical knowledge in teachers;
- Teachers have the basic language, subject and pedagogical skills to absorb the new knowledge and skills available from TDP;
- The materials are appropriate for teachers working with students of different abilities, particularly around language; and
- There are no changes to the curriculum or other features of the education system that render these materials redundant.

82. Overall, the evidence shows that not all teacher training interventions have been successful in improving teacher effectiveness, and that the structure and quality of the teacher training is of critical importance (Westbrook, 2013). Westbrook's systematic review of teacher training programmes in developing countries found that teacher education can best support effective pedagogy when: a) teachers receive peer support; b) professional development is aligned with teachers' needs, the promoted pedagogy and modes of assessment of their practice and follow-up monitoring of teachers; c) there is support from head teachers; and d) forms of assessment are aligned with the curriculum.

83. The TDP in-service programme, with its use of direct training, continuous monitoring, peer support, head teacher support, and alignment with the curriculum, has been designed to avoid common pitfalls of cascade teacher training. However, it is likely that in northern Nigeria there may be challenges with both the current capacity of teachers, head teachers and teacher educators to absorb new skills, and the management environment that encourages them to participate in training and monitoring. For example, TDNAs conducted in northern Nigeria have highlighted significant deficiencies in some teachers' subject knowledge and language skills which may limit their ability to absorb training (Johnson and Hsieh 2014). In addition, available evidence summarised in Humphreys and Crawford (2014) does not indicate that all schools and LGEAs have positive management environments. It is currently unclear how the TDP will

respond in schools where management is weak, or where head teachers select inappropriate teachers for training.

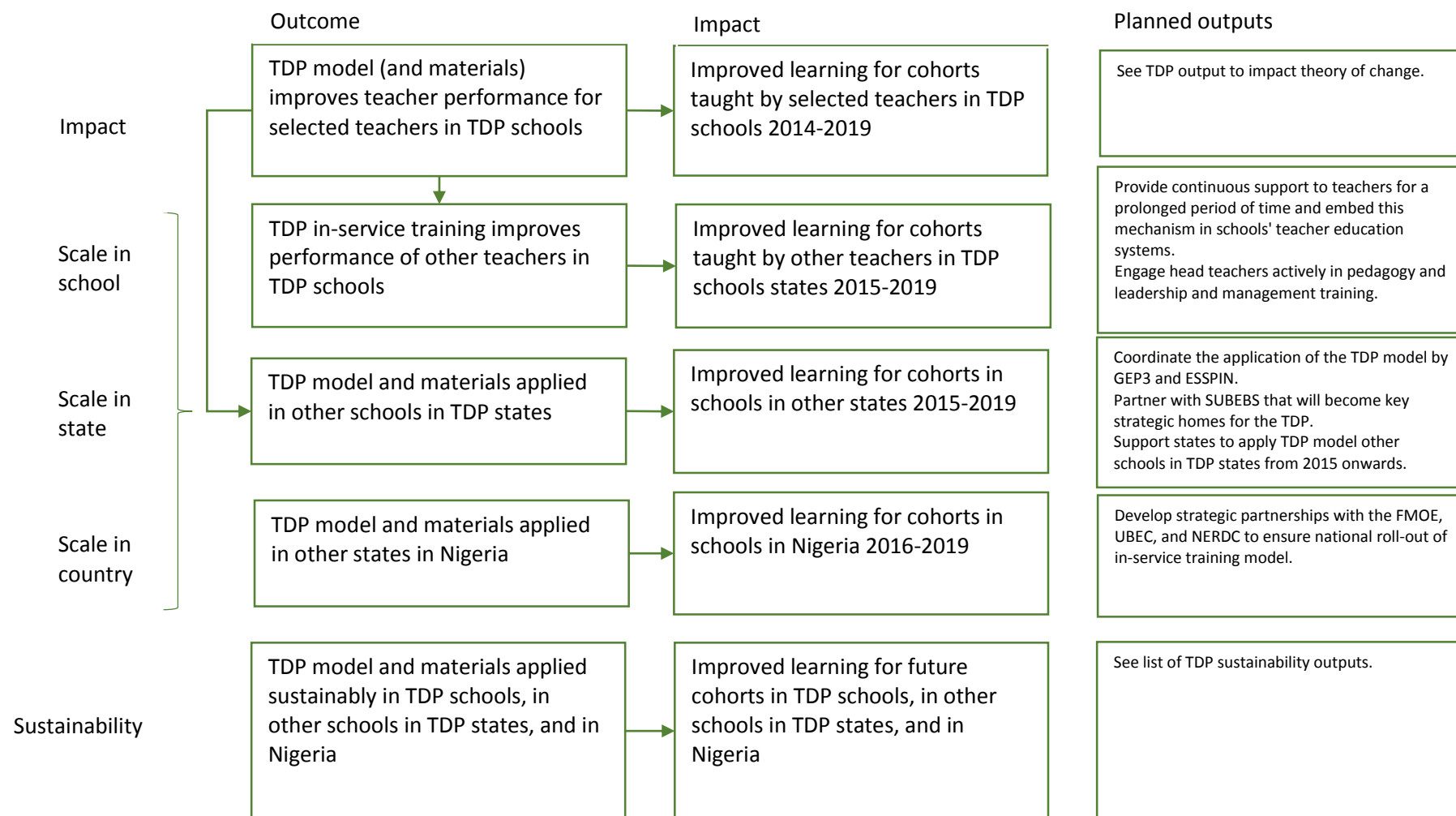
84. The cascade model presents some risks. Reviews of the effectiveness of cascade training (see e.g. Wilson and Smith 2013) have indicated that the benefits of training can be diluted by the time they reach the classroom unless there are very significant inputs to training at each level. While the details of this training modality are still to be worked out fully, it is not obvious that the current level of inputs will be sufficient to avoid this pitfall. In addition, the support given to teachers will be dependent on the skills and application of individuals in the Teacher Development and Teacher Facilitation teams, and these skills will vary. Synergies with the TDP's pre-service component may partly address 'training readiness' for some individuals, but this will take time to feed through into the current in-service rounds of training.

2.1.2.4 Scale and sustainability

85. While the primary objective of the TDP is to improve learning outcomes for children attending schools in which TDP in-service activities take place, the programme also seeks to improve teaching and learning in other schools, and has a range of activities and strategies intended to achieve this. Specifically the TDP aims to move beyond this primary impact by encouraging the application of the TDP model (including materials) to:

- To other teachers in TDP-supported schools,
- To other schools in TDP states,
- To other schools in Nigeria, and
- In the years after the TDP ends in 2019.

86. In Figure 5, we set out a summary of these additional scale and sustainability impacts, and list some of the outputs mentioned against these impacts. These outputs are largely part of the **collaboration and partnership** part of the TDP in-service strategy. Below, we develop these outputs and the assumptions behind them in more detail.

Figure 5: How the TDP aims to achieve scale and sustainability

Scale in schools

87. The TDP does not provide training directly to all teachers in the schools in which the programme operates (only up to four including the head teacher), but it envisages that the benefits of the model will apply to other teachers in these schools. The strategies most intended to support this in-school scaling are:

- Provide continuous support to teachers for a prolonged period of time and embed this mechanism in schools' teacher education systems.
- Engage head teachers actively in pedagogy and leadership and management training.

88. Most of the assumptions between TDP outputs and intermediate impact set out above are relevant for this sort of scaling, but there are also some additional key assumptions that could be tested:

Table 3: Assumptions for TDP achieving scale in TDP schools		
Outputs	Assumptions	Outcomes/impact
TDP model working to improve teacher effectiveness	<ul style="list-style-type: none"> • Head teachers apply TDP approaches to other teachers, including meetings, support, and monitoring 	<ul style="list-style-type: none"> • TDP in-service training improves performance of other teachers in TDP schools
Continuous support model embedded in TDP school systems	<ul style="list-style-type: none"> • Other teachers are capable, motivated and managed to adopt TDP approaches with more limited support 	<ul style="list-style-type: none"> • Improved learning for cohorts taught by other teachers in TDP schools states 2015-2019
Head teachers actively engaged	<ul style="list-style-type: none"> • TDP materials are shared with other teachers 	

89. There is no good source of evidence on these assumptions currently; they would need to be assessed as part of an evaluation (on which more below). However, in general the capacity development literature is cautious about the likelihood of these sorts of spontaneous transfers of capacity without strong leadership and effective management.

Scale in states

90. The TDP in-service strategy is designed to improve the chances of the TDP model being applied in other schools in TDP states. The key activities related to this (excluding those related to results and evidence) are:

- Coordinate the application of the TDP model by GEP3 and ESSPIN,
- Partner with SUBEBS that will become key strategic homes for the TDP, and
- Support states to apply TDP model other schools in TDP states from 2015 onwards.

91. The TDP's interactions with GEP3 and ESSPIN are relatively easy to manage because each programme is funded by DFID, and ESSPIN is implemented by a sister company to that implementing the TDP. The TDP's use of existing ESSPIN and GEP3 approaches to head teacher training and some classroom materials will also help this. However, as both programmes have limited timeframes (ESSPIN until 2017; GEP3 until 2020), the ability of these programmes to engender long-term change to teacher in-service training practice in schools with which they work may be limited if this is not a core activity with the same level of support given by the TDP.

92. Collaboration and partnership with SUBEBs is a key part of the TDP's scaling strategy. The TDP judges that the CoEs in their states are not yet ready to implement in-service teacher training, and that since SUBEBs are currently responsible for teacher training as part of basic education, this is the key

institutional home for the TDP in-service training. This judgement seems sensible, but is of course vulnerable to changes in the CoE reform agenda, and to the capacity of SUBEBs in each state. The establishment of permanent teacher educator teams using existing government school support structures should also improve the chances of the TDP model being applied in other parts of the state. The state-specific TDP in-service strategies will provide more detail on how states will be supported to expand teacher support in other schools from 2015 onwards, but this will of course be dependent on the interest and capacity (including funds) of state institutions. The TDP have also approached the development of materials used in the training in such a way to try to ensure that they are relevant to the national curriculum and expertise, in order to support the scaling of the model (including materials) to other schools and states.

93. Major assumptions for achieving scale in TDP states are set out below. Again, the evidence base for these assumptions needs developing.

Table 4: Assumptions for TDP achieving scale in TDP states		
Outputs	Assumptions	Outcomes/impact
Coordinate the application of the TDP model by GEP3 and ESSPIN,	<ul style="list-style-type: none"> ESSPIN and GEP3 have the interest and capacity (managerial, technical and financial) to apply TDP model 	<ul style="list-style-type: none"> TDP in-service training improves performance of teachers in other schools in TDP states
Partner with SUBEBs that will become key strategic homes for the TDP, and	<ul style="list-style-type: none"> ESSPIN and GEP3 can apply TDP model successfully in their timeframe 	<ul style="list-style-type: none"> Improved learning for cohorts in schools in other states 2015-2019
Support states to apply TDP model other schools in TDP states from 2015 onwards.	<ul style="list-style-type: none"> SUBEBs remain the appropriate institutional home for in-service teacher training State governments/SUBEBs consider in-service teacher training a priority, and the TDP model appropriate, on a sustained basis SUBEBs have the capacity (managerial, technical and financial) to apply TDP model in other schools 	

Scale countrywide

94. The TDP also has ambitions for its in-service programme to be rolled out across Nigeria (though again, this is not an explicit target). This is to be achieved principally through the output:

- Develop strategic partnerships with the FMOE, UBEC, and NERDC to ensure national roll-out of in-service training model.

95. While this is not an explicit target for the TDP, in the context of DFID's support overall this is an important ambition for the programme. The modalities for these partnerships have been set out by the TDP, but it is clear that they will need to evolve against the background of the changing political landscape, including after the February 2015 gubernatorial and national elections, and the changing policy and fiscal environment in Nigeria. There are of course a large set of assumptions around the national roll-out of the TDP model, but as many of these overlap the sustainability assumptions we have included in the section below.

Sustainability

96. As with most development projects, the most difficult and most important goal is the sustainability of TDP outcomes and impact after project funding ceases in 2019, both in TDP schools, and in other schools in Nigeria. The TDP has some strategic approaches designed to ensure this, some of which also contribute to scale (and to other outcomes). These include:

- Develop strategic partnerships with the FMOE, UBEC, and NERDC to ensure long-term sustainability of in-service training model;
- Partner with SUBEBS that will become key strategic homes for the TDP;
- Ensure state contributions for procurement and management of technology devices;
- Engage with the EDRCs at state level for teacher training material development and strengthen its capacity to become the custodian of all audio-visual resources within and beyond the project period;
- Produce audio-visual material on SD cards;
- Engage local experts in developing materials;
- Recruit a pool of teacher educators as teacher development team;
- Recruit teacher facilitators from the current school supervisor cadre;
- Embed mechanism of continuous teacher support in schools and state teacher education systems; and
- Engage head teachers actively in pedagogy and leadership and management training

97. These outputs are expected to provide support to the implementation of the TDP in-service model, allow the model to be implemented outside the TDP states, and to improve the model's long-term sustainability. These are important goals, but there are various assumptions that would be required to hold for this to occur, including that:

- these institutions are the appropriate institutional homes for an in-service teacher training model,
- the TDP as an externally funded time-bound programme has the capacity to develop these partnerships,
- these institutions are open to these partnerships being developed, on a continuous basis (e.g. in a way that is resilient to changes in their leadership and political standing),
- these institutions have the capacity (broadly defined to include e.g. funding, human resources, management capacity, skills, etc.) to engage in the in-service training model,
- reform to teacher effectiveness remains an appropriate policy priority to improve learning in Nigeria,
- reform to in-service teacher training remains an appropriate policy priority to improve teacher effectiveness in Nigeria,
- a single in-service model is appropriate for all different states in Nigeria (rather than different models in different states),
- the TDP in-service model is appropriate in Nigeria, and improves teacher effectiveness and learning,
- training materials can be updated to reflect changes to the curriculum, language of instruction, etc.,
- the technology remains robust to technological and infrastructural change and can be maintained in working order.
- the same local experts continue to be involved in teacher training,
- the SUBEBs and LGEAs have the incentives and capacity (managerial, financial, technical) to maintain, support and renew the teacher educator teams without TDP support, and
- head teachers and teachers have the interest and capacity to continue to support the TDP in-service training model.

98. The validity of these assumptions about the future is difficult to pre-judge and there will probably be major changes to the Nigerian education system, politics and financing that impinge on many of these. However, evidence on some of these will become more apparent during the programme.

2.2 Conclusion to the theory of change

99. This section on the TDP's in-service theory of change is designed to inform the development of the evaluation framework, specific questions and therefore approaches to the evaluation. We have expanded on the documentation provided by the TDP to include more detail on the expected causal links and assumptions between them. This will allow the testing of the programme theory. As more detail on the TDP's approach is developed, this theory of change can be further enhanced.

3 Evaluation design and framework

We first set some context for the evaluation framework. Primary users are DFID Nigeria and the TDP, and secondary users include Nigerian government institutions. The budget and timing of the evaluation are set out, followed by the TDP Results and Evidence Strategy, DFID criteria for deciding what to evaluate, and then a list of evaluation questions. We propose a theory-based approach to the evaluation of the TDP's in-service component, which aims to make a credible judgement of whether the theory of change holds in practice, and why. This requires that we make the theory of change explicit (as we did in section 2), including assumptions about causal links (how the TDP aims to bring about change), the programme implementation (how the TDP intervenes to set change in motion), and assumptions about external context (what conditions must be in place to make the causal chain and implementation work).

3.1 Evaluation users

100. The primary users of the in-service evaluation are DFID Nigeria and the TDP. The evaluation should therefore be designed with the needs of these groups in mind, and they should both be consulted as far as possible in designing the evaluation (with signoff coming from DFIDN and TDP, and ideally the TDP's National Coordinating Committee, and being involved through extensive communication in the conduct of the evaluation (attending presentations, seeing designs, observing training, receiving focused dissemination, etc.).

101. The secondary users of the evaluation include 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. These groups are important because they should apply the results of the evaluation directly in their in-service activities. The evaluation communication should therefore be structured around these users.

102. Additional audiences for the evaluation include LGEA staff, head teachers and teachers, as well as students and their parents, the DFID Research and Evidence Division (RED), other education donors in Nigeria, and other education projects in Nigeria. These additional audience groups should be closely involved in dissemination activities. In particular, TDP may have opportunities to maximise impact by involving headteachers, teachers, students and parents in the evaluation, because they will be able to recognise that TDP activities have a positive impact on teacher and student performance, which should improve their commitment to the programme. Although beyond the scope of the evaluation team's work, we recommend that TDP engage these stakeholders in dissemination activities, supported by the evaluation team and EDOREN's knowledge management workstream.

103. The evaluation will also be of use to other researchers and policymakers involved in improving education systems in Nigeria and elsewhere. These secondary audiences will be invited to participate in dissemination activities, but it is not proposed to design specific communications around them (beyond a minimal academic publication and conferencing from the evaluation).

3.2 Evaluation resources

104. TDP have provided 25 days for evaluation design, and this has been supplemented by EDOREN resources, particularly around some of the highly complex sampling and evaluation design issues. TDP have a budget of around GBP1.1 million for results and evidence, and GBP700,000 of this is available for an evaluation baseline. DFID have an additional GBP1.7 million for evaluation, so the total evaluation budget available is GBP2.4 million.

105. The major evaluation expense will be surveys. A survey in three states on teaching and learning that is designed to detect a small change in learning outcomes at State-level (a sample of 336 schools, 1344

teachers and 2,688 students) is likely to cost around GBP700,000, including design, survey, analysis and reporting. The midline survey (on teachers only) is expected to cost around GBP600,000 and the endline GBP700,000, making a total of GBP2 million for evaluation surveys, and leaving around GBP400,000 for qualitative fieldwork.

106. This implies that three large teaching and learning surveys are possible within the current budget. It is therefore unlikely that both phase one and phase two can be comprehensively evaluated using quantitative methods to assess both learning and teaching, as this would require four very significant surveys at minimum, and DFID have requested that this does not take place.

107. The preferred evaluation survey design, considering financial and other constraints is to focus on the TDP phase one states, with a minimal proposal of a learning and teaching baseline in 2014, teaching follow up in 2015, and learning and teaching follow up in 2017. This leaves additional resources for other activities related to other aspects of the theory-based evaluation that do not involve large surveys (such as the qualitative research set out in this framework).

108. Two other options were considered and rejected:

- i. The evaluation could have assessed teaching change only in the TDP phase one states, with a baseline in 2014 and follow up in 2015 just on teachers. The TDP activities in phase two states could be evaluated for both teaching and learning, with surveys in 2016 and 2018. This would not meet DFID's need for a learning baseline in 2014, and would not give much time for impact on learning to obtain in phase two states (which run 2016-2019).
- ii. The evaluation could have assessed teaching and learning in the TDP phase 1 states, with surveys in 2014 (teaching and learning), 2015 (teaching only) and 2017 (teaching and learning), and teaching only in the TDP phase 2 states, with surveys on teaching in 2016 and 2018. This option would have generated more data but would have been too costly.

109. The rest of the evaluation costs are for the qualitative fieldwork. For the three rounds of fieldwork in three states and four schools per state (see section 4.2.1), this is estimated to come to GBP390,000 in total.

110. The total costs of the evaluation are therefore GBP2,390,000, which is within the total evaluation budget of GBP2.4 million. The baseline quantitative survey will be contracted through the TDP's budget, and the remainder directly contracted by DFID.

3.3 Evaluation timing

111. TDP runs from 2013 to 2019, with the inception phase ending in 2014 and implementation beginning in Jigawa, Katsina and Zamfara in 2014. In-service support to teachers in these states starts in year 1 (school year 2014/15) with teachers teaching primary grades 1-3, and then extends to training on materials for primary grades 4-6 from SY 2015/16, and JSS from SY2016/17. Roll out to the next three states (Kaduna, Kano and Niger) will take place from September 2016, for in-service across all grades. EDOREN's first phase runs from 2013 to 2017, and a potential second phase from 2017 to 2021.

112. Given the importance of formative evaluation, formative evaluation results will need to be produced in time to feed into design work for the TDP phase two in early 2016. This should include a survey of changes in teacher pedagogical competence and an explanation of which parts of the programme are working and why, but not an assessment of learning outcomes. These results will apply to the in-service component.

113. Given that JSS activities do not start until SY2016/17, the evaluation will focus on TDP activities in grades P1-6.

114. The learning aspects of the evaluation (i.e. has the TDP had an impact on student learning and why?) will be presented later, as it is usually the case that changes in learning outcomes take longer to obtain. This implies final results on learning from the TDP phase one in 2018/2019, which would take place in EDOREN's second phase.

115. Given that these results will need to be generated in part from surveys, and these are the least flexible components of an evaluation (qualitative research can be conducted more flexibly), it seems reasonable to propose an ambitious survey calendar at this point:

Table 5: TDP in-service evaluation quantitative fieldwork timing			
Survey timing	Location	Phase	Purpose: evaluation of the impact of in-service training on:
October 2014	Phase 1 states (Jigawa, Katsina and Zamfara)	Baseline	Teacher quality and learning achievement in English, maths and science&technology for TDP-selected teachers and a panel of pupils starting p3 (tested on the p2 curriculum)
October 2015	Phase 1 states (Jigawa, Katsina and Zamfara)	Midline	Teacher quality in English, maths and science&technology for TDP-selected teachers, to feed into TDP phase 2 design.
October 2017	Phase 1 states (Jigawa, Katsina and Zamfara)	Endline	Teacher quality and learning achievement in English, maths and science&technology for TDP-selected teachers and the same panel of pupils now starting p6 (tested on the p5 curriculum)

116. The timelines for the reports are therefore:

- March 2015: Baseline quantitative report
- June 2015: Baseline report including quantitative and qualitative data.
- March 2016: Midline quantitative report
- June 2016: Midline report including quantitative and qualitative data.
- June 2018: Final report including quantitative and qualitative data.

117. If additional budget is available, surveys on the TDP phase 2 could take place, and follow up surveys to check sustainability of changes would be a possibility in 2020, which would be very interesting as a measure of whether positive impact is sustained when TDP is no longer providing support. However, given that resources for this are not confirmed, these follow ups are not included in this framework.

3.4 The TDP's Research and Evaluation (R&E) Strategy and logical framework

118. The TDP evaluation proposed here is congruent with the TDP R&E strategy and logical framework (provided in full in Annex C), although it will not duplicate it. The indicators and data sources for the proposed evaluation draw on those proposed in the R&E strategy and logical framework.

119. The evaluation framework proposes to use these suggested measures and sources for output level data for the in-service evaluation, without suggesting any changes to the logframe or the approach in the R&E strategy, but will extend these outputs as per the discussion in section 2.1 on the theory of change. It is expected that the activities to measure these outputs will be conducted by the TDP, with technical support from EDOREN where necessary.

3.5 DFID five key criteria for deciding what to evaluate

120. DFID provide key criteria for deciding what to evaluate, and we use these to focus the evaluation.

Table 6: DFID Five Key Criteria for deciding what to evaluate

No.	Criterion
1	How strong is the evidence base supporting the intervention design? Interventions with a weak evidence base, perhaps because they are innovative, are likely to be a priority for evaluation to learn lessons for future programming.
2	Is the intervention designed as a pilot , with a plan to scale up or transport to a different context if successful? If so, then clear evidence on how the intervention works is likely to be required.
3	Is the intervention contentious? In cases where there are different views about the likely success of an intervention it may make sense to evaluate to understand what does happen and be transparent in this.
4	What do stakeholders think? It may make sense to evaluate when an intervention is of wider strategic interest, where there are opportunities to work with partners – including helping build capacity in evaluation or where stakeholders have a clear preference for an evaluation. Where programmes are being undertaken jointly, it would usually be expected that any evaluation would be a joint undertaking as well.
5	Financial value – all large value projects (i.e. above £5m) should at least be considered for evaluation, but it may not be cost effective to evaluate lower value projects if there are no other reasons for doing so.

Source: DFID (2012) Evaluation Handbook

121. **How strong is the evidence base:** As described in the TDP’s business case, overall the evidence for the theory of change can be summarised as being strong for the links from outcome to impact, but more limited and emerging for the link from inputs to outputs, and outputs to outcomes. The detailed discussion of the theory of change above in Section 2 highlighted a number of links for which evidence in northern Nigeria is currently weak, at all stages of the causal chain. We did not come across any study which robustly established, within a single study, the links between teacher training, increased teacher quality, and higher learning outcomes. The limitation of the evidence is noted in the recent systematic review on teaching in developing countries (Westbrook et al 2013).

122. In addition, TDP’s delivery of training through the use of ICT (“trainer-in-pocket”) comes from another DFID education programme in Bangladesh called English in Action. Though a number of studies were conducted to assess the effect of EiA on various impacts, none were studies designed to attribute causal impact. In addition, it is difficult to comment on how easily transplantable the EiA model is for the context of Nigerian schools and teachers.

123. **Is the intervention designed as a pilot:** As mentioned above in Section 1.3, both DFID and TDP see significant formative value in TDP’s evaluation, particularly its in-service component. Specifically, the evaluation should test in TDP phase one whether the in-service model for improving teacher competence works overall, which parts work well, which parts work less well, and why. The purpose of this part of the evaluation is to inform phase two. This means that the evaluation should test changes in teacher competence that result from in-service work, and provide initial explanations for why changes are occurring, with evidence available by early 2016 (meaning data collection in 2015, or TDP year 2).

124. **Is the intervention contentious?** As such the TDP intervention is not contentious. Its design borrows heavily from DFID Bangladesh's English in Action (EiA) programme which, even though it did not produce robust causal impacts on learning outcomes, showed benefits for teacher effectiveness and learning outcomes, and did not raise any ex-ante or ex-post concerns for contentious views on outcomes and impact. There are some outstanding concerns around whether trained teachers would leave the teaching profession for better opportunities in non-teaching jobs, or leave rural schools for postings in urban settings, etc. but these would need to be monitored by the programme for design lessons.

125. **What do stakeholders think?** As per TDP's inception report and activities in the first six months, stakeholders across various federal, state and LGA levels see strategic and long-term value in investing in teacher training. This has also been recognised by the FMOE as an important area for improving teacher quality and learning outcomes. Given time constraints, local stakeholders (parents, teachers, students) have not been consulted in the development of this framework.

126. **Financial value:** TDP is a £53.6 million project with its in-service component bearing 80% of the impact weighting (£37.2 million from DFID, and another £16 million approx. from counterpart government funding).

3.6 Evaluation questions

127. Evaluation questions were developed based on both the discussion of priorities, resources and timing above, and technical considerations of what sort of measurement is possible. As we present in section 4.1 below where we set out the rigorous impact evaluation design, discussions with DFID and TDP indicated that there was scope to ask and answer questions around quantitative changes in teacher effectiveness and learning outcomes that are attributable to the TDP.

128. Evaluation questions are structured around the DAC evaluation criteria and the detailed theory of change set out above. A list of evaluation questions was developed during the preparatory phase of designing the TDP evaluation framework and have been refined subsequently. The evaluation matrix presented in Annex G sets out questions by evaluation criteria (relevance, effectiveness, etc.), the area of the theory of change to which they pertain (linked to the discussion above), and presents indicators, analytical approaches, how information will be collected to answer the question, and the level of disaggregation. Not all assumptions or causal links have been included; some have been removed because they are not critical to the evaluation purpose or it is not feasible to answer the questions given the resources available to the evaluation.

129. The evaluation questions in Annex G are structured around relevance, efficiency, effectiveness, impact and sustainability, and questions in each of these areas are organised around the components of the theory of change developed above. Questions pertaining to relevance are largely structured around the design and assumptions of the TDP. Questions on efficiency relate principally to value for money and the efficiency of implementation – linking to process evaluation questions. Questions on effectiveness, impact and sustainability are principally concerned with those aspects of the theory of change that relate to outcome and impact areas, and the assumptions that underpin them. The questions have focused on the areas of the theory of change where evidence for assumptions is weakest, where outcomes are most important, and where there is largest scope for learning as per the evaluation purpose and the DFID criteria for what to evaluate. A summary of these questions is given in the table below:

Table 7: Key TDP evaluation questions		
OECD DAC Evaluation criteria	Key Question	Formative/summative and timing of answer
Relevance	Are TDP's assumptions correct globally and particularly for the Nigerian education and policy context?	Formative - 2016

Relevance	Does TDP's in-service training approach and design address needs, priorities and constraints of the primary teachers in northern Nigeria?	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

3.7 Theory based evaluation

130. The evaluation approach combines pragmatism with aspiration while acknowledging the presence of budget constraints. At the same time, the evaluation options are designed to promote learning for other contexts, while emphasising the importance of contextual analysis and being able to form argumentative links (Cartwright and Hardie 2012).

131. The evaluation is based on the TDP's theory of change – the sequence and composition of the TDP activities – the results chain - expected to lead to the TDP's desired outcomes. The TDP ToC has been discussed extensively above. In this section, we set out our approach to how the theory-based evaluation of in-service training will be operationalised. The next section discusses the evaluation plan in more detail.

132. While the attribution approach uses experimental or quasi-experimental methods to generate rigorous evidence on impact (as set out below), the theory-based contribution analysis will assess the TDP by following its theory of change and gathering data on the key assumptions and context as well as expected outputs and outcomes. In this way, following e.g. Cartwright and Hardie (2012), the evaluation will be able to argue that the TDP has led to certain outcomes, provided there is evidence on the assumptions, context and outputs.

133. The theory of change disaggregates the results chain while highlighting each step on the ladder to TDP impacts: *'improved student learning outcomes and amount of schooling'*. The final step before impact is improved teacher effectiveness (effective new teachers and effective existing teachers). The in-service component seeks to improve the subject knowledge and pedagogical competence of existing teachers, primarily through school-based cascade training, materials, support, the trainer in the pocket, and continuous support to teachers.

134. In addition to the rigorous quantitative information gathered from school clusters in the six states where TDP in-service activities are taking place (and controls), as suggested in the TDP R&E strategy (McCormick 2013) the TDP evaluation will also use mixed methods to conduct what White (2009) calls 'rigorous factual analysis' to examine whether the expected links in the causal chain hold.

135. This theory-based analysis asks: 'Why did this impact occur? Did this activity lead to this output, to this outcome, to this impact?', and where the assumptions are valid, over time (i.e. Is the context as we assumed it was?)'

136. Theory-based evaluation of the TDP will follow six principles (White 2009):

- a. Map out the TDP's causal chain, as has been done in section 1.2;
- b. Understand context, as has been done in section 1.2;
- c. Anticipate heterogeneity;
- d. Rigorous evaluation of impact using a credible counterfactual;
- e. Rigorous factual analysis. This will be applied across the TDP's causal chain; and
- f. Use mixed methods. The impact evaluation surveys (of children and schools) will be combined with qualitative instruments, in an iterative way.

137. In accordance with the principles of Contribution Analysis we will also give serious considerations to the possible alternative, non-project explanations for the changes observed (see Kotvojs and Shrimpton 2007).

3.7.1 Map out the TDP's causal chain

138. Mapping out the TDP's causal chain and context (in section 2.1) allowed the development of an evaluation matrix and specific evaluation questions and indicators (see Annex G) about whether the TDP is achieving its outputs, outcomes and impact and which contextual factors are required for this. Coupled with some secondary data analysis of the context in other areas of Nigeria to check whether these contextual assumptions hold there as well, this will allow some speculation about the likely impact of the TDP if implemented at scale. This generalisation is not perfect, because it cannot account fully for the existence of other education programmes in these districts, but the process of generalising will take these programmes into account as context, based on available secondary documentation on these programmes.

3.7.2 Understand the TDP's context

139. The identification of context that is relevant to the TDP theory of change, and therefore allows the possibility of generalising results, is not straightforward and various approaches to this have been proposed in the literature. In essence, these approaches share a careful analysis of what would have to hold if the impact of a programme found in one place was to apply in another context: 'thinking horizontally and vertically' (Cartwright and Hardie 2012). The identification of relevant context is therefore a theoretical exercise, informed by research available on similar programmes and on TDP. The development of this context will be an iterative process (White 2009), whereby relevant aspects of context are identified in fieldwork and review, and then investigated further, through the most appropriate qualitative means (for instance document review or primary fieldwork at different levels).

140. A first step to the identification of the TDP's context has been attempted in section 1.1 above, but much more detail will be required to be able to describe in detail whether the TDP's teacher-training focused approach to improving education outcomes will work in different contexts. Specifically, this will require checking the contextual assumptions that have been made in the theory of change and that are critical to impact (for example, that training improves the effectiveness of teachers in Nigeria, even without better or good teacher deployment or management). This is a detailed process that is set out to a degree below, but will require an iterative approach (i.e. developing new contextual issues to assess at each stage of fieldwork) to ensure all relevant context is taken into account.

3.7.3 Is the TDP going to have heterogeneous impact?

141. TDP impacts are likely to be heterogeneous to the extent that different teachers and schools will be more responsive to interventions than others. The direction of impact on children taught by trained

teachers is likely to be positive. However, as discussed above, if improved teacher effectiveness attracts new children to classes, the effect on *average* learning in a class may not be positive, if new students have lower learning levels.

142. How might this be dealt with? Ideally, the survey could sample children who enter school during the TDP implementation as well as children already in school, and differentiate between the two. However, this would be costly and it is not obvious that a sufficiently number of children would join school to be able to detect an impact on their learning outcomes. The proposed impact evaluation survey will therefore sample from children already in school, to provide an estimate of the TDP impact on children already in school.

3.7.4 Can we establish the TDP's impact using a credible counterfactual?

143. The rigorous evaluation of impact using a credible counterfactual focuses on the in-service training and will tell us whether learning improvements, if they occur, can be credibly attributed to this output. This is achieved through constrained randomisation and discussed in section 0.

3.7.5 Rigorous factual analysis across the TDP's causal chain

144. Rigorous factual analysis will be conducted on the TDP in-service component, and within these along the most significant causal chains for eventual impact. This, together with the qualitative methodology proposed, is explained in more detail in section 4.2. The baseline quantitative and qualitative fieldwork will help further focus subsequent rounds of qualitative data collection on specific issues in the context and on causal pathways that are considered by stakeholders and researchers to be most significant in terms of either their contribution to overall impact or their uncertainty. To this extent, the evaluation will be amenable to a certain amount of speculation about what would happen if the two different components of the TDP were scaled up. It will also help address the larger policy question around whether the TDP model would work elsewhere in Nigeria. If the evaluation were to find that the TDP worked in the programme states, how would policymakers know whether it might work elsewhere? While there are at present no agreed best practice for how to answer this question, the impact evaluation use the theory of change for this purpose (Cartwright and Hardie 2012).

145. Rigorous factual analysis along the causal chain will allow the evaluation to play a formative role. Changes in learning outcomes take time to appear, but the theory-based evaluation will be able to gather qualitative and quantitative information in the shorter-term on whether TDP activities are leading to improved teacher training and effectiveness. These early outcome level data (from the in-service quantitative surveys in 2015, as well as qualitative data) will subsequently be complemented with impact level data on whether greater effectiveness has led to improved learning (in the 2017 endline).

146. This causal chain analysis will complement the rigorous counterfactual impact assessment. For the in-service output, programme monitoring data on training conducted etc., secondary data (e.g. on teacher numbers and deployment), and data from in-depth interviews and focus groups, primarily with government officials, teachers and students, will complement the detailed information that comes from the quantitative surveys.

3.7.6 Use mixed methods

147. The case for a mixed methods approach to impact evaluation is strong and several tools for mixing methods are available (White 2009), and suggestions for using different types of data are also outlined in the TDP R&E strategy. There are three main considerations: 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). There are eight proposed steps to integrate methodologies

for better measurement; six suggested steps to sequence information for better analysis; and the final step is to merge findings at analysis stage into a single, mixed methods report (Garbarino and Holland 2009).

148. Mixed methods designs are well suited to evaluate complex programmes like the TDP that cover multiple levels of the system (working with teachers, schools, training institutions, and government system level engagement). The evaluation plan proposes using both quantitative surveys and qualitative approaches and the evaluation will therefore draw on different sources of data. This is important for several reasons. First, 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 the types of explanations that can be drawn around these data. Second, different data sources can help to triangulate findings. Third, the 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 informs and is combined with the analysis of the other. Finally, qualitative methodologies can be more flexible to ongoing changes to the programme and therefore 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 quantitative survey to determine impact, and qualitative data will play a stronger role in the assessment.

149. Specifically, the TDP evaluation follows these practical steps proposed by Garbarino and Holland (2009).

- i. Use indicators from the baseline quantitative surveys to help to select a qualitative investigation sample;
- ii. Use baseline quantitative survey results to highlight priority issues and generate new hypotheses to cover in qualitative research;
- iii. Use qualitative analysis (the reviews of the literature and stakeholder consultations conducted in developing this evaluation framework) to identify knowledge gaps to be filled by the quantitative surveys;
- iv. Use qualitative analysis (the reviews of the literature and stakeholder consultations conducted in developing this evaluation framework) to prioritise issues important to stakeholders that should be covered by the quantitative surveys;
- v. Use qualitative analysis (the reviews of the literature and stakeholder consultations conducted in developing this evaluation framework) to construct indicators;
- vi. Use qualitative analysis (the reviews of the literature and stakeholder consultations conducted in developing this evaluation framework) to define sub-group sampling frames;
- vii. Use a qualitative study in a sub-sample of quantitative areas to compare findings;
- viii. Use a qualitative study to assess heterogeneity behind quantitative averages from the surveys;
- ix. Use the qualitative study to explain relationships emerging from the surveys;
- x. Use the qualitative study to triangulate survey findings;
- xi. Use the qualitative study to enrich analysis of relationships/trends/patterns from the surveys;
- xii. Produce integrated mixed methods evaluation reports that draw on both quantitative and qualitative data.

150. The TDP evaluation will follow a positivist approach to evaluation that sets out to test certain pre-developed or emergent hypotheses using these different data sources (rather than a constructivist approach that allows research participants to construct meaning around the TDP's activities and impacts). In order to facilitate testing both pre-developed (this framework) and emergent hypotheses, the research proposed around the evaluation framework will take an iterative approach, where the findings from the baseline will inform the questions asked at the midlines, and so on to the endlines. This iterative process will also work by learning from the TDP phase 1 for the evaluation of the TDP phase 2. This is important

because it is possible that TDP's approach changes in phase 2 as a result of emergent findings in early 2016 of the evaluation of phase 1 (the evaluation's formative role). But it is also important because any evaluation should respond to unanticipated findings and seek to explore them if they are important.

151. Taken together, these steps generate an evaluation plan that addresses the questions and measures the indicators set out in Annex G, which also sets out the analytical approaches, information collection methods, and levels of disaggregation for each question. The next section on the evaluation plan provide more detail on the analytical approaches and data collection methods.

4 Evaluation plan: proposed fieldwork

This theory-based evaluation will include a rigorous assessment of whether the TDP's in-service model works to improve teacher effectiveness and learning outcomes, using a (constrained) randomised control trial. It also includes tests, though mixed methods, of other parts of the causal chain and assumptions from output to impact at scale. Finally, it includes a process evaluation to verify that the TDP is achieving its outputs as intended.

4.1 Rigorous impact evaluation design

152. Given the importance of the in-service component of the TDP for learning, a rigorous impact evaluation is designed around this output, using a quantitative survey. The quantitative impact evaluation will focus on assessing the impact of TDP in-service training support on teacher effectiveness and pupils' learning outcomes.

153. The quantitative impact evaluation design proposed here is resource-intensive, primarily due to the extensive survey data collection required. The key cost driver for quantitative impact evaluation studies is generally whether or not the evaluation can take advantage of pre-existing survey or administrative data that covers TDP schools and includes information on the key impact indicators under study. In this case it has been established that no such pre-existing data is available, so there is no alternative except to conduct a large-scale survey to measure teacher knowledge and pupil learning outcomes in TDP schools.

154. Furthermore, the requirement for a very robust impact evaluation methodology means that this survey needs to be conducted at three points in time. The TDP phase 1 states: at 'baseline', just before the TDP programme activities begin; and at 'endline', which is proposed to be three years after the baseline survey, but also a midline one year in, to provide formative information to the TDP phase 2. The reason for repeated survey rounds is so impact indicators can be compared 'before' and 'after' the TDP programme began operating, therefore enabling the specific changes that took place over the first three years of TDP operations to be measured. The robust impact evaluation design also affects the scope of the survey in another way. In order to assess whether the changes in impact indicators in TDP schools between baseline and endline are due entirely to TDP (and not other prevailing factors, such as generally improving school standards), the survey also covers non-TDP schools, teachers and pupils which act as a comparison group. In other words, the evaluation survey has to cover both TDP and non-TDP schools at baseline and endline.

4.1.1 Research questions

155. The quantitative surveys have been designed specifically to address the following major evaluation questions:

- What is the impact of the TDP in-service activities on teacher effectiveness in the classroom? (*evaluation question Effe-1*)
- What is the impact of the TDP on teacher absenteeism? (*evaluation question Effe-2*)
- What is the impact of the TDP in-service activities on teacher subject pedagogical knowledge? (*evaluation question Effe-4*)
- What is the impact of the TDP in-service activities on pupils' learning outcomes? (*evaluation question Im-1*)

156. However, while these four questions underpin the design of the quantitative surveys in terms of sampling and approach, these surveys will also produce data on several other evaluation questions, as set out in the table in Annex G.

4.1.2 Measuring learning outcomes

157. The TDP impact is anticipated to be on student learning outcomes. However, there are various complexities in operationalising this indicator, of which some but not all are discussed in the R&E strategy. These include:

- What is language of testing for student learning? The ESSPIN MLA, on which the TDP MLA is proposed to be based, asks literacy and numeracy questions in English, as do the teacher quality instruments. This is likely to underestimate levels of learning overall, but provided that the test can differentiate between low levels of learning, this doesn't matter because the evaluation is designed to assess change in learning outcomes rather than the levels of learning. The ESSPIN instruments are being improved to better assess low levels of learning.
- Does a change in student learning outcomes of 3% mean that average scores on a test go up by 3 percentage points or that 3% more students reach a certain benchmark on a test? Is student learning calculated using raw scores or item response theory? The evaluation framework is currently designed around measuring a change in average raw scores – i.e. the target is interpreted as a 3 percentage point increase in average raw scores of students taking the test. However, the evaluation team will continue to investigate the possibility of using item response theory to establish a more nuanced assessment of student aptitude, and/or the possibility of linking results (whether raw scores or aptitude measured using IRT) to certain criteria of achievement expected at their grade.
- Should the same students be tested at different times of their school career (a panel), or should learning change be measured by repeated cross-sections of students at particular grade levels (such as at p2 or p4)? The current evaluation framework proposes using a panel (of pupils at the start of P3 at baseline), because this allows the detection of smaller changes for a given sample (even accounting for sample attrition, which is estimated at 62.5% to account for dropout at higher primary grades).
- What is a reasonable time lag between the beginning of the teacher training intervention and the emergence of improvements, if any, in learning outcomes? The current evaluation framework attempts to leave as long a gap as possible between starting the intervention and measuring learning outcomes – given the use of a panel this will be three years.
- MLA tools for the evaluation will need to be carefully designed and tested so as not to avoid jeopardising comparability over time and across the different education programmes and surveys that make up DFID's Nigeria programmes.
- Since teachers being trained may not belong to a specific grade, which grades of pupils should the evaluation focus on for measuring learning outcomes? In practice, teachers could be a) teaching a particular subject across grades; or b) teaching all subjects to a particular grade or class; or c) teaching all subjects to a particular cohort. Given the variability in practice across different schools (Bennell 2014, draft), the evaluation framework proposes to use a cohort/panel approach to follow students as they move through TDP teachers. However, the exposure of this cohort to TDP-trained teachers will be monitored in pre-tests to check the validity of this approach.

4.1.3 Overall approach

158. One of the key challenges in designing the type of robust impact study required for this evaluation, is defining a suitable comparison or 'control' group. Ideally, the evolving characteristics of the control group should provide a reliable counterfactual for what the 'treatment' group – i.e. those benefiting from the programme – would have experienced in the absence of the programme 'intervention'. By comparing the actual experience of the treatment group to its hypothetical counterfactual (as captured by the experience

of the control group), it is possible to accurately attribute the various changes measured as programme impact.

159. Since a fully randomised design approach is not feasible (in particular the need to allocate all schools into clusters), the impact evaluation will instead be based on a quasi-experimental ‘constrained randomisation’ approach.⁷ This approach was recommended in the TDP R&E strategy. The programme was originally planning to ask SUBEBs to select one cluster of schools per LGEA to be covered by the programme. .

160. The SUBEBs selected schools on the following basis:

- In Jigawa, since nearly all schools (by September 2014) will have exposure to ESSPIN, the TDP in-service working group (consisting of SUBEBs, LGEA officials, TDP state staff, among others) selected two sets of 12 school in 14 LGAs by trying to establish a balance in the level of exposure to ESSPIN across the two sets. Each set of schools was clustered to allow for regular cluster-meetings and further peer learning among the trained-teachers. Generally speaking, two ESSPIN clusters (each of which have 5-7 schools in them) composed a single TDP cluster of 12 schools.
- In Katsina and Zamfara where there are no pre-existing clusters of schools (these are not ESSPIN states), the in-service working group will similarly select 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 GEP3 states, a programme which is undergoing a phase of re-design. The current idea is for the programmes to liaise to ensure GEP3 schools are selected from outside the TDP treatment and control clusters.

161. In order to facilitate the impact evaluation, the programme is now requiring SUBEBs to select two school clusters per LGA. It is important that the two clusters within each LGA are broadly similar, but this cannot be checked ex ante because the quality of existing data on schools and teachers is too low. The guidelines provided to the SUBEBs to assist them in this are shown in Annex F.

162. Within each LGA, schools in both clusters will then be required to select the four teachers who will potentially benefit from the programme (referred to as ‘potential TDP teachers’). The programme’s criteria for selecting teachers is as follows: the headteacher in each school is always selected (whether they teach or not), plus another three teachers selected by the headteacher. However, in practice it is not clear that these criteria are necessarily being correctly applied by the SUBEBs. This isn’t a problem from an evaluation perspective, because we would anticipate that the adherence to the teacher selection criteria would be similar across treatment and control schools.

163. Once the teacher selection has been made in all schools across both clusters, one of the clusters will be randomly selected to receive ‘treatment’ (i.e. be covered by the programme) with the other being allocated as ‘control’ (i.e. will not be covered by TDP within the evaluation period). This random allocation of the treatment cluster could be undertaken by the evaluation team. However, it is instead recommended, for reasons of transparency, for representatives of the SUBEBs and both clusters to be present at the random allocation (e.g. done by tossing a coin, or using a scratch-card) in each LGA. The evaluation team recommends that control schools should be prioritised to benefit from TDP (or other programmes such as ESSPIN or GEP3), once the impact evaluation has been completed (i.e. after October 2017).

⁷ In this context a fully randomised experimental evaluation of TDP would require all public primary schools in the states in which TDP operates to be allocated into clusters. These clusters would be allocated in line with TDP’s plans for defining its clusters of TDP schools. Within each state a large number of clusters would be randomly selected to be covered by the programme (treatment clusters), with a similar number randomly selected as the comparison group (control clusters). Crucially, because the selection of the four TDP teachers within each school does not follow any set criteria, the schools’ identification of the TDP teachers would have to occur after the treatment and control clusters have been selected, but before they have themselves been informed of their treatment status.

164. The schools in the treatment clusters are the treatment population; the schools in the control clusters are the control population. A representative sample of treatment and control schools, treatment and control teachers, and treatment and control pupils will then be drawn to be covered by the evaluation survey. The sampling design is outlined in section 4.1.6.

165. It is important that treatment and control schools have an equal degree of overlap with ESSPIN and GEP3. In the ESSPIN state (Jigawa) the TDP school clusters will be based on the ESSPIN clustering of schools. In other words, in Jigawa for each cluster pair, both clusters within the pair will be based on ESSPIN clusters and only contain ESSPIN schools. Therefore in Jigawa all treatment and control schools will be ESSPIN schools. In the case of the two GEP states (Zamfara and Katsina), the proposal is that the TDP will not work in GEP3 areas, and vice versa. It is important that any overlapping programmes (currently ESSPIN) will operate identically in treatment and control schools. It should be noted that this design means the evaluation will not be able to assess whether TDP works better in combination with ESSPIN, because it won't be possible to distinguish this from state-level variations. Instead it will provide estimates of the relative impact of TDP plus ESSPIN compared with TDP alone in Jigawa.

166. If possible, the evaluation team will try to assess the degree to which SUBEBs have succeeded in selecting similar cluster pairs within each LGA by collecting key information on every school within every selected cluster. This would ideally include information on location (rural/urban), school size, number of teachers per school, average class size, pupil-age mix, average pupil performance (if available) and whether any other programmes (e.g. ESSPIN,) are in operation. Such cases can be followed up (through phone-calls to SUBEBs or even site visits), to verify whether the cluster is genuinely poorly matched. Where poorly matched clusters exist, the evaluation team would have the option of excluding outlier schools which are skewing the match, from the treatment and/or control school sample frame prior to drawing the evaluation sample. In other words, even if there are some poorly matched cluster pairs, the sample frame can be adjusted to try and minimise the impact these apparent mismatched cases might have on skewing the comparability of our treatment and control samples. Section 4.1.7 below discusses how the similarity (or 'balance') of the treatment and control schools, teachers and pupils will be separately verified using the baseline survey data, and, should statistically significant differences exist, what can be done to address this in the analysis.

167. Under this proposed evaluation design, four 'potential' TDP teachers will be chosen in all schools across every selected pair of clusters in each of the 14 TDP LGAs in each state. After the random allocation of the programme within each pair of clusters, only half of these 'potential' TDP teachers (i.e. those in treatments schools) will actually go on to benefit from the programme. The other half (those in the control schools) will not be covered by TDP. It is therefore very important that this process is clearly and transparently communicated to headteachers in advance of the teacher selection, so that they fully understand that going through the teacher selection process does not mean the school will actually end up being covered by the programme. It is also suggested that headteachers should not tell selected teachers about their 'potential' TDP status until after the random allocation has been made and it is confirmed that their school will be covered by TDP, so as to avoid causing disappointment for control teachers – i.e. selected teachers that ultimately will not be covered by TDP.

168. In some cases, it may be that teachers that are listed as selected by the TDP are not found to be in schools when the evaluation fieldwork is conducted (e.g. because they have been transferred). In these cases, the TDP and the evaluation have agreed that there will be a replacement teacher selected for the TDP programme by the evaluation fieldwork team. This replacement will be done in the school in discussion with the headteacher and the TDP team (by phone), and the replacement will be interviewed.

169. There is a possibility that some SUBEBs and/or headteachers may want to indirectly compensate the schools and teachers that end up as controls by choosing them to benefit from other opportunities that might arise. If this practice is common, and these other alternative interventions have similar impacts to TDP, this could potentially undermine the evaluation by 'confounding' the impact estimates. In other words, because the impact is measured by comparing the improvement made in TDP schools with those in

non-TDP treatment schools, if many of the non-TDP treatment schools are receiving alternative interventions which create a similar level of improvement, there will be no discernible difference between the TDP and non-TDP schools on average, and therefore the impact estimates will suggest the programme had no impact when in fact it did. It is not feasible (or ethical) to restrict the scope of non-TDP control schools to benefit from alternative programmes, so this is a genuine risk. It is therefore extremely important that the endline survey captures information about alternative programmes working with control schools and teachers. This information can then be used to assess the prevalence of these potential confounding factors, and if necessary to try and adjust for them in the analysis.

170. There is also an option of the programme directly compensating control teachers in some way, for example by providing them with a one-off training day, so they at least feel they have had some benefit from the process. However, the risk here is that if this compensation exercise is effective in improving teaching quality and pupil outcomes this may ‘confound’ the impact estimates, in the same way to that described in the previous paragraph. However, it is very unlikely that a single, one-off training day would have sufficient impact to have this confounding effect – so the option of direct compensation of control teachers probably shouldn’t be ruled out on technical grounds (but it may be financially unfeasible). But in any case, provided headteachers are made aware of the process (and the possibility that selection of potential TDP teachers does not necessarily mean the school will benefit from the programme), with selected teachers ideally not informed of their status prior to confirmation of the programme allocation, then there is no obvious need to directly compensate control teachers.

4.1.4 Timing of survey rounds

171. The sequencing and timing of the evaluation survey rounds has to be planned around TDP roll-out and the academic terms: for TDP phase 1 states (Jigawa, Zamfara and Katsina):

- Baseline 1 (Oct 2014): capture baseline levels for p2 English, maths and science&technology learning attainment for students starting their p3 year, and teacher effectiveness for English, maths and science&technology for TDP-selected teachers.
- Midline (Oct 2015): capture midline levels of teacher effectiveness for English, maths and science&technology for TDP-selected teachers.
- Endline (Oct 2017): capture endline levels for p5 English, maths and science&technology learning attainment for the same cohort of students, now starting their p6 year, and teacher effectiveness for English, maths and science&technology for TDP-selected teachers.

4.1.5 Instruments

172. The impact indicators listed in Annex G will be measured using a range of survey instruments:

- A head-teacher interview (one per school): this will capture attitudes, motivations and school-level characteristics that may influence our impact indicators (so we can control for these in the impact analysis if required), including head teacher training and exposure of the school to other interventions such as ESSPIN and GEP, again so that we can control for this in impact analysis; headteacher supervision of teachers; and records of teacher absenteeism during the past five school days.
- A teacher interview (all four TDP teachers per school): this will cover attitudes, motivation and teacher-level characteristics that may influence our impact indicators (so we can control for these in the impact analysis if required), including previous training they have received, perceptions about teacher-teacher interaction, absence from school and reasons for absence, supervision by

the headteacher, use of mobile phone. Most of these questions will also be integrated into the headteacher questionnaire, but only addressed to headteachers who teach class.

- Classroom observation: one class [N1] observed for the first class available for each of the four TDP teachers per school, or three teachers if the headteacher does not teach: this will be used to assess the teaching skills of each TDP teacher in the schools.
- Teacher subject and pedagogical knowledge test (all four TDP teachers per school): this will take place as part of the school survey.
- Combined English, maths and science & technology tests to assess learning improvements between p2 and p5 (eight pupils per school): for the October 2014 baseline this will be a combined p2-level test to be conducted on a sample of pupils taught by at least one TDP/control teacher at the start of their p3 school year; at endline (Oct 2017) these same pupils will be at the start of p6 and will be given a combined p5-level test (plus a limited number of p2-level questions to maintain direct comparability with the baseline).

173. The procedures for sampling the schools, and selecting the TDP pupils to be tested, are set out in section 4.1.6 below.

4.1.6 Survey sampling design

174. The sample design is determined to a large extent by practical programme considerations, and also by the available budget. Within these parameters the design is intended to maximise the statistical power of the impact indicator 'difference-in-difference' measures (and minimise the Minimum Detectable Effects).

175. The programme will operate in 14 clusters per state. As set out above there will also be an additional 14 control clusters in which schools will go through the TDP selection process but where the programme will not operate. Clusters are expected to comprise of around 12 schools on average. In each cluster four schools will be randomly selected for the evaluation survey (that is, about one third of the treatment and control schools). This gives a total of 112 schools (56 treatment and 56 control) per state, and 336 across all three phase 1 TDP states.

176. In every sampled school, the headteacher plus the three other TDP/control teachers will be interviewed. Each TDP/control teacher will also have one class observed and, following the completion of the survey, they will also be given a subject knowledge test, which will take place at a testing centre once the school survey is complete. Each round of the evaluation survey will therefore cover 448 teachers (224 TDP and 224 control) per state, and 1344 in total for the three states.

177. These same teachers will be re-interviewed, observed and tested in the 2015 midline and endline surveys so improvements in their teaching skills and subject knowledge can be measured. It is anticipated that the programme will only support TDP teachers while they remain at their original school. If a teacher moves school, they will cease to be covered by the programme, and a new TDP teacher will be selected in their place. Therefore only TDP/control teachers that remain in their original school will be re-interviewed. In other words, the survey will not trace TDP/control teachers who move schools.⁸ Note that some bias may emerge if TDP training affects the 'business-as-usual' attrition of teachers. In other words, teachers benefiting from the TDP school may be less likely to move specifically because they don't want to stop receiving TDP training, and therefore the teacher turnover is lower among TDP teachers compared to the

⁸ There is an option to interview the replacement TDP teachers, but as replacement of control teachers is unlikely to be feasible in control schools, these replacement TDP teachers would not be included in the impact analysis, so the only purpose would be for assessing the replacement teachers' perspectives and attitudes towards the programme.

control teachers, If attrition is high it would be useful to track a sample of both TDP and control teachers that their schools and assess how different they are.

178. In order to assess pupil learning outcomes, for the initial baseline survey eight of all the pupils starting p3 in September 2014 and who will be taught English, maths or science & technology by at least one TDP/control teacher will be randomly selected to be tested on their knowledge of p2 material.⁹ This random selection will be made in-field, using CAPI or paper-based sampling, drawn from a 'sampling frame' of all the p3 pupils that are being taught these subjects by the TDP/control teachers for that school year and who are present in school on the day of the survey. This gives a sample of 896 pupils per state (448 treatment and 448 control), and 2688 in total.

179. In the event that in some schools TDP/control teachers teach no p3 pupils it is proposed that the entire school is replaced. This approach is justified on the basis that our focus is on the impact of TDP on pupil outcomes as measured by the improvement of the p3 cohort. There is however a chance that it is only the more disorganised schools where TDP teachers have been selected such that none teaches any p3 pupils (which is unexpected given the TDP teacher selection criteria). This could potentially introduce a bias (albeit one that is balanced across treatment and control schools), meaning that our impact results are not exactly representative of the TDP programme overall. However, the evaluation team feel this potential risk is outweighed by the need to maintain the treatment and control pupil cohort sample sizes.

180. The same pupils who were tested at baseline will be re-interviewed in the endline survey. This will enable us to measure the impact of TDP on improving performance between the start of p3 and the start of p6 for the cohort of TDP pupils who had just completed p2 and were about to start p3 at the baseline (October 2014). This approach is referred to as 'panelling pupils'. While the pupils will be given a p2-level test at baseline, they will be given a p5-level test at endline when they are beginning p6 in October 2017 (plus a limited number of p2-level questions to maintain direct comparability with the baseline).

181. The cohort approach implies no testing at p4 and no repeated testing at p2. This will save money for the evaluation by allowing the detection of impact at lower cost. This is not likely to be problematic because ESSPIN and GEP3 will probably generate repeated estimates of learning at p2 and p4 levels in their evaluation surveys, so doing this again for the TDP evaluation would not add much to this. However, if the TDP or DFID require additional rounds of p2, we have calculated the minimum detectable effects for impact using a repeated cross-section (at p2), and this approach with the current sample would be able to detect impacts on average p2 test scores for English, Maths and Science&Technology above 4.2-4.8%. This would relate to the average impact across all three states. For individual state-level analysis the minimum detectable impact would be changes in pupils' p2 test scores of at least 7.3-8.3%. If the follow up sample were increased, these effect sizes would come down slightly (i.e. a smaller impact would be detectable). A repeated cross-section may end up being a sensible approach to take depending on how the programme tends to operate in practice. If 'exposure' to TDP teachers is focussed in the early years (i.e. p1-p3) then it may be the case that our cohort of sampled pupils, who are starting p3 in September 2014, may only end up being taught by a TDP teacher for the first year of the evaluation period (p3, 2014-15). In this case the limited exposure to the programme among our intended cohort could mean that our cohort-based estimates of pupils' p2-p5 test score improvements may suggest the programme is having limited or no significant impact. It may be, however, that the programme does have a significant impact on p1 and p2 learning which would be picked up by repeatedly testing repeated cross-sections of pupils at the start of p3 (using a p2 test). The evaluation will monitor pupils' exposure at the midline and make a decision on whether a repeated cross-section would make more sense.

182. This survey sampling strategy is summarised in the table below:

⁹ The rationale for administering combined subject tests to all sampled pupils is both pragmatic and technical. So the risk is that some of our supposedly 'treated' maths pupils may actually only be taught by a TDP teacher for English, and vice versa.

Table 8 Sample design summary

	Treatment	Control
Evaluation clusters per state	14	14
Sampled schools per cluster	4	4
Headteacher interviews per school (baseline ; midline; endline)	1	1
Teacher interviews/observations/tests per school (baseline ; midline; endline)	4	4
Pupils tested for English and Maths improvement between p2 and p5 per school (baseline 1; endline)	8	8

183. The intended sample sizes for each survey round (corresponding to the survey sampling strategy) are shown in the table below. Note that due to attrition the final sample sizes to be used for the impact analysis may be considerably lower.

Table 9 Sample sizes (by survey round)

	Treatment (per state)	Control (per state)	Total (per state)	Total (all three states)
Baseline 1 (October 2014)				
Headteacher interviews	56	56	112	336
TDP/control teacher interviews	224	224	448	1344
TDP/control teacher classroom observations	224	224	448	1344
Total teacher English, maths and S&T subject and pedagogical knowledge tests	224	224	448	1344
Total pupils (Oct14 p3 cohort) tested: p2-level combined English and maths	448	448	896	2688
Midline (October 2015)				
Headteacher interviews	56	56	112	336
TDP/control teacher interviews	224	224	448	1344
TDP/control teacher classroom observations	224	224	448	1344

Total teacher English, maths and S&T subject and pedagogical knowledge tests (midline)	224	224	448	1344
Endline (October 2017)				
Headteacher interviews	56	56	112	336
TDP/control teacher interviews	224	224	448	1344
TDP/control teacher classroom observations	224	224	448	1344
Total teacher English, maths and S&T subject and pedagogical knowledge tests	224	224	448	1344
Total pupils (Oct14 p3 cohort, now in p6) tested: p5-level combined English and maths	448	448	896	2688

184. This sampling design is associated with specific Minimum Detectable Effects (MDEs) for each of the indicators listed above. These are shown in the tables below. They indicate the minimum level of effect that the programme needs to have on each of the impact indicators in order for this effect to be measured as being statistically significant using the proposed survey design.

185. Details of how the MDEs have been calculated, along with key underlying assumptions, are set out in C.1. The tables below, and the annex, include estimates for pupils at p4, because this was considered important for the TDP. However, we propose not to test pupils at p4 because we propose instead a more cost effective cohort approach.

Table 10 Minimum detectable effects: state-level impact indicators

Impact area	Indicator	Subject/detail	Baseline average (percentage point score)	MDE (percentage points)	Final sample size (treatment)	Final sample size (control)
Teacher knowledge	Increase in average p1-3 TDNA assessment score	English	8.1	5.1	112	112
		Maths and S&T	47.5	5.5	112	112
	Increase in average p4-6 TDNA assessment score	English	8.1	5.1	112	112
		Maths	47.5	5.5	112	112
		Science & Technology	47.5	5.5	112	112
Teaching skills	spent in groups or pairs	20.0	3.7	112	112

	Increase in average proportion of lesson time... teacher involves students in positive interaction	20.0	3.7	112	112
Pupil learning outcomes	Increase in average test scores: p2 vs p5 (Oct14 p3 TDP pupil cohort)	English	35.3	6.5	168	168
		Maths and S&T	52.1	5.7	168	168

Notes: (1) Sources for baseline means and standard deviations for teacher-level indicators: Teacher knowledge: GEP3 Katsina TDNAs; Teaching skills: guesstimated (in the absence of other information) by EDOREN as at baseline 20% of lesson time is spent in pairs and for 20% of lesson time the teacher involves students in positive interaction; Pupil learning outcomes: ESSPIN Composite Survey 2. (2) Final sample sizes based on conservative assumed attrition rates: 50% for teachers (between baseline and endline); 62.5% for the Oct14 p3 pupil cohort. (3) The intra-cluster correlation (ICC) is assumed to be 0.3 for teachers and pupils. (4) The inter-temporal correlation (ITC) is assumed to be 0.8 for teachers and pupils.

Table 11 Minimum detectable effects: state average impact across all three TDP states

Impact area	Indicator	Subject/detail	Baseline average (percentage point score)	MDE (percentage points)	Final sample size (treatment)	Final sample size (control)
Teacher knowledge	Increase in average p1-3 TDNA assessment score	English	8.1	2.9	336	336
		Maths and S&T	47.5	3.2	336	336
	Increase in average p4-6 TDNA assessment score	English	8.1	2.9	336	336
		Maths	47.5	3.2	336	336
		Science & Technology	47.5	3.2	336	336
Teaching skills	Increase in average proportion of lesson time...spent in groups or pairs	20.0	2.1	336	336
	 teacher involves students in positive interaction	20.0	2.1	336	336
Pupil learning outcomes	Increase in average test scores: p2 vs p5 (Oct14 p3 TDP pupil cohort)	English	35.3	3.7	504	504
		Maths and S&T	52.1	3.3	504	504

Notes: See Table 5 above.

4.1.7 Internal validity

186. Our evaluation design is relying on the randomisation of treatment within pairs of similar TDP clusters to deliver treatment and control populations that are similar in their observable characteristics. In this case, our control population represents a valid counterfactual for what the treatment population would have experienced in the absence of the programme.

187. Therefore after the completion of the baseline survey we will test for statistically significant differences between the treatment and control populations (schools, TDP teachers and TDP pupils). If, by chance, the randomisation process has failed to deliver balanced treatment and control populations (i.e. we observe some statistically significant differences in their baseline characteristics) then this will have to be addressed in the analysis. The diff-in-diff approach will by design account for baseline differences, but we may also want to measure (in the endline survey) and control for time-varying factors that are likely to co-influence the impact indicators. This is because if our treatment and control populations exhibit differences in their static baseline characteristics then we may also expect differences in the degree to which they are affected by time-varying factors (i.e. violation of the so-called ‘common trends’ assumption). Another approach that we could consider as a back-up option is to use matching techniques

(eg PSM) to identify sub-groups within our treatment and control samples that are sufficiently comparable; however, this may be challenging in the context of the clustering of schools. Our current understanding is that the SUBEBs are following the randomisation process.

188. There is also a risk that some of the treatment pupils only receive limited or partial ‘treatment’, both in terms of the length of time and/or the number of subjects they are in contact with a TDP teacher. For example, although a sampled treatment pupil will be taught by at least one TDP teacher in the initial academic year by design, there is no guarantee that they will continue to be taught by TDP teachers in subsequent academic years. Similarly, not all sampled treatment pupils will be taught by TDP teachers for all three subjects (English, maths and science&technology), but they will be tested on all three. In other words, it is likely that some proportion of our treatment sample will in fact have only had limited exposure to TDP teachers. If this turns out to be very common then the programme impact effects that we measure will be in some sense ‘diluted’. Whether this is seen as problem or not depends on whether we are more primarily interested in the impact of the programme on all pupils with *any* exposure to the programme, as opposed to the impact on all pupils with full or high exposure to TDP. Our default approach will be to estimate the former, but if the prevalence of partial treatment is very high, then we can also estimate the specific impact on those with full (or at least very high) exposure, by restricting the analysis to exclude pupils with partial exposure. The problem here is that this results in reducing the sample sizes, and therefore the size of impacts that can be detected. It also depends on the midline and endline surveys capturing information for all the treatment and control pupils on who they were taught by for each subject in each year, and who among these were TDP teachers. However, on the other hand, this variation in treatment ‘exposure’ can actually be exploited to add another dimension to the impact analysis: we can test whether treated pupils with greater treatment exposure (i.e. those who received more teaching from TDP teachers over the evaluation period) experience greater improvements in their learning outcomes compared to treated pupils receiving lower ‘exposure’.

189. Another threat to internal validity is differential attrition: this occurs if treatment teachers and/or pupils are significantly more or less likely to disappear from the sample (i.e. be untraceable at endline) compared to controls. For example, we may anticipate that attrition will be lower for treatment teachers because they will only be covered by the TDP while they remain at the TDP school, and therefore have this additional reason to remain in their job. On the other hand, TDP may strengthen their teaching skills to such an extent that they are more likely to be offered other better jobs at other schools, which would be associated with relatively higher attrition. The extent of differential attrition among treatment and control teachers and pupils will be measured in the endline survey and, if significant, efforts will be made to account for it in the impact analysis.

190. Since treatment and control schools will be clustered within the same LGAs, there may be a chance that teachers will be transferred between treatment and control schools. This potentially presents another threat to internal validity because some of the control pupils may end up being taught by teachers who have been partially treated. However, this is unlikely to occur often enough to be a concern.¹⁰

191. If teachers are compensated, either directly or indirectly, this would also pose a threat to internal validity. As discussed above, direct compensation is unlikely to be necessary, but indirect compensation (e.g. in the form of other projects like GEP3 implementing TDP-like interventions in the control clusters) may occur and it would be impractical and unethical to prevent it. Therefore attempts will be made to assess this in the endline survey (i.e. by asking control teachers what alternative programmes and initiatives they have benefitted from), and to account for it in the analysis.

192. A final threat to internal validity may arise if the TDP teachers and/or pupils know that their competencies will be compared with a control group. There is a risk that such knowledge may change their

¹⁰ If we wanted to monitor this, we could include in the endline headteacher interview questions about whether any new teachers coming into the school were covered by TDP at their previous school, and then cross-check this against the sampled pupils’ teachers.

behaviour, i.e. ‘raise their game’ in order to compare favourably. And it is also possible that the control group will act in a similar way. Provided this effect plays out equally on average among the treatment and control groups then this won’t pose a problem for the analysis, but if the incidence varies between treatment and control this could potentially confound or exaggerate the impact results. The problem is that it is impossible to observe this effect, and therefore to assess its extent and control for it. Instead the best approach is to ensure treatment and control teachers and pupils receive exactly the same information about the purpose of the evaluation survey, and have identical understanding about the context of the assessments they will undertake. Ideally both groups should be told the assessments are being undertaken to measure the ‘typical’ (rather than individual) performance of teachers over time, with no explicit mention of any comparison with another set of teachers/pupils, and that the testing cannot be traced back to them and that there are no “stakes” associated with it.

4.1.8 External validity

193. The theory-based evaluation is designed to help answer questions about the TDP that can be applied in other contexts, given the emphasis in the evaluation of learning. The qualitative research set out below and recommended by the TDP R&E strategy will help to answer additional questions about why impacts occurred. As such, the rigorous impact evaluation is only a component of the evaluation approach – albeit the component with the most significant resource implications.

194. The impact evaluation survey results will be representative of the treatment and control populations. However, these evaluation populations are not themselves representative of the three phase 1 TDP states more broadly. This is because the TDP clusters are being chosen by SUBEBs in each LGA, rather than being randomly sampled from a comprehensive list of potential TDP clusters (into which all schools have been allocated). This has important implications for the external validity of the impact evaluation findings.

195. For this reason it will be important to understand (through analysis and by consulting stakeholders) how the TDP schools differ from other schools in Jigawa, Zamfara and Katsina in general. This will enable judgements to be made as to whether the impact evaluation results are a good indication of the expected impact if TDP were scaled up in these states. In turn, assessments will be made as to whether the impact evaluation results can be reliably generalised to the phase 2 TDP states as well as other states in Nigeria more generally.

4.1.9 Phase 2 TDP states: Kano, Kaduna and Niger

196. The above evaluation design only relates to the phase 1 TDP states (Jigawa, Zamfara, and Katsina). To measure impact in the three phase 2 TDP states (Kano, Kaduna and Niger) would require at least two additional rounds of survey fieldwork (baseline and endline). Given available resources, it is currently not proposed to conduct an evaluation on phase 2 states. If additional budget is made available, the TDP phase 2 states would be assessed only in terms of the TDP’s in-service activities’ impact on teacher effectiveness, not on student learning.

4.2 Rigorous factual analysis

197. The remainder of the evaluation questions will be answered using rigorous factual analysis based on the TDP’s theory of change, and a process evaluation designed to test whether the TDP is implemented as designed and planned. In this section, we will not repeat the discussion on the theory of change (section 2.1) or the evaluation questions (Annex G), but will link questions to methodologies and discuss methodologies in more detail. The methodologies overall will follow the approaches proposed in the TDP R&E strategy.

198. The rigorous factual analysis will draw on the following data sources:

- The quantitative surveys of teacher training, teacher effectiveness (including classroom observations) and learning outcomes in TDP in-service schools in phase1 states in 2014, 2015, and 2017 that have been discussed in the section above.
- TDP monitoring data directly collected by TDP from schools on teacher attendance and training activities conducted.
- TDP programme records on training activities conducted
- Qualitative evaluation fieldwork in 2014/15, 2015/16, 2016/17 and 2017/18 (managed by EDOREN, but potentially also TDP). This will be particularly important for answering several of the evaluation questions listed above. This is explored in more detail in section 4.2.1.

199. Additional secondary sources (e.g. evaluations of other programmes, ESSPIN Composite survey data, GEP3 baseline data, teacher supply and demand studies, economic and other relevant contextual data from northern Nigeria, etc.).

4.2.1 Qualitative methods

200. The qualitative fieldwork methodology is both confirmatory (testing existing hypotheses) and exploratory (explaining impacts, developing new hypotheses and capturing unexpected impacts). Broadly, fieldwork with children, parents, teachers, headteachers, teacher facilitators and teacher educators, LGEA and SUBEB staff, and DFID, TDP and national government staff, is proposed in 2015, 2016, 2017 and 2018. The methods used will principally include key informant interviews with some focus group discussions, conducted by EDOREN researchers. Qualitative fieldwork will follow the quantitative surveys so that initial findings from the quantitative surveys can be used to develop detailed qualitative interview schedules. Analysis of the qualitative findings will be conducted alongside the quantitative analysis to produce single integrated and mixed methods reports.

201. To answer the evaluation questions linked to qualitative research, the qualitative research will involve a series of case studies of students and teachers in TDP treatment schools, to be tracked throughout the programme to explain any changes in effectiveness and learning outcomes, as well as interviews with other (non-TDP) teachers and selected fieldwork in non-TDP schools in the control group to assess spillovers. The evaluation questions to be answered by qualitative research are set out in Annex G. These questions include both impact evaluation questions (i.e. questions on the translation from TDP outputs to outcome and impact) and process evaluation questions (i.e. questions on the TDP's achievement of expected outputs).

4.2.1.1 Rigour

202. A major methodological challenge in qualitative research is the definition and achievement of 'rigour', particularly, as here, when the research methodology should be open to the identification of new hypotheses, causes and unexpected impacts. Qualitative research is often accused of being 1) open to research bias or anecdotal impressions, 2) impossible to reproduce and 3) difficult to generalise (Mays and Pope 1995). The sections below briefly set out the approach to addressing these questions, in terms of sampling, fieldwork, and analysis.

4.2.1.2 Sampling

203. Methodological rigour in qualitative research is not best established through a statistically representative sample, because results cannot be quantified and aggregated in the same way as quantitative data can be. But as in quantitative research, rigour in qualitative research can be achieved

through “systematic and self-conscious research design, data collection, interpretation and communication,” (Mays and Pope 1995: 110).

204. As with most qualitative research, the chosen approach to sampling is theoretically informed and designed to generate responses from small numbers of individuals and groups that are representative (though not statistically) of groups relevant to the TDP, and which allow some identification of heterogeneous impact. Specifically, purposive ‘typical case sampling’ (at the highest level, sampling an average school in terms of education outcomes) and ‘extreme case sampling’ (sampling high and low performing schools in terms of selected education inputs, outputs and outcomes) will be used. This is not designed to produce results that are generalisable in the same sense as quantitative data. Rather, the generalisability of the qualitative research results derives from the extent to which they are embedded in a theory of change that has some validity in a wider context. This form of sampling allows exploration of what the TDP is doing in a typical case, but also performance in ‘high impact on teacher effectiveness’ and ‘low impact on teacher effectiveness’ cases. At this stage we propose a tentative sampling design, to be reviewed in the context of the quantitative baseline when complete.

205. How large a sample is appropriate? Probably the most effective way to set a sample in qualitative research is to do this flexibly, when theoretical saturation is reached – when no new data, categories or relationships seem to be emerging (Cataldo et al. 2011). In the present context, there is need to be more specific, and there are resource limitations. As such, the sample will be linked to the number and types of individuals that will be concerned with the TDP and related to its impact, and attempt to provide a reasonable spectrum of teacher training outcomes to inform the sampling.

4.2.1.3 Fieldwork

206. Rigour and the avoidance of bias in the qualitative fieldwork will be achieved through extensive training, and the involvement of different individuals in the field teams, so that the teams can provide checks on each other. In addition, the teams will keep records of their activities, so that they can be linked to the transcripts and analysis.

4.2.1.4 Analysis

207. The analysis for each evaluation question to be answered in part or in full by qualitative research will use inference. Rigour in the analysis of the qualitative data comes from four principal sources. First, findings will be triangulated against different data sources, both qualitative and quantitative. Second, different members of the team will conduct and discuss the analysis, reducing the possibility of individual researcher bias. Third, the analysis sheet will be made available externally, with confidentiality controls. Finally, the analysis will be subject to peer review.

4.2.1.5 Structured and unstructured methodologies

208. The first methodological choice is across the continuum between more structured methods and more unstructured methods (Cataldo et al 2011). Almost all qualitative methods are less structured than those used in a survey, where only a limited set of responses are possible. Since here it is necessary to both confirm and explore hypotheses, and to be sensitive to new ones, both relatively structured and relatively unstructured methods will be used.

209. Structured qualitative methods in this context means having a set of specific questions (e.g. ‘have you observed any changes to teachers’ pedagogical practices? What are they? Why did these changes take place’) that prompt discussions in a focus group or are put to interviewees, or which researchers answer through their observations. These interview questions are generated first from the evaluation questions set out above, and second from analysis (to be performed once data are collected) of the baseline quantitative

data that will generate hypotheses to be tested in qualitative research. The evaluation indicators and respondents in the table below will generate an initial set of qualitative interview schedules.

- The advantage of these structured methods is that they permit research teams to be specific and therefore efficient in testing pre-specified hypotheses. This is appropriate to the extent that the research will test these hypotheses that have been developed and have a certain amount of contextual knowledge of the setting through the research teams, a review of literature, and initial stakeholder interviews and field tests.
- The disadvantage is that they can leave little space for respondents to introduce unexpected ideas or explanations and are often not appropriate for addressing potentially sensitive topics (such as low teacher aptitude).

210. Unstructured qualitative methods in this context means having a set of areas to probe, but without specific hypotheses around them. In practice, this would mean having much looser prompts in interviews or focus groups (at their loosest: ‘have you seen any changes in the way education is provided in these classes, if so what?’), and looser checklists for researchers to observe. These prompts are based on less clearly specified hypotheses (for instance areas that stakeholders suggested exploring) or on no hypothesis at all. The qualitative approaches will also include unstructured questions.

- The advantage here of an unstructured approach is that the research teams can capture impacts or explanations that were not anticipated, or aspects of the context that were not initially considered important but turn out to be. This is the only methodology within the impact evaluation that permits this. This is appropriate for evaluation questions on unanticipated impact (*evaluation question Im-3*)
- The disadvantage is that an unstructured approach can be inefficient, as respondents may describe changes or explanations that are not directly related to TDP.

4.2.2 Proposed qualitative fieldwork

211. The evaluation will draw on qualitative fieldwork in 2015, 2016, and 2017. At the school/child/parental level, fieldwork will be carried out in the treatment LGAs where the quantitative fieldwork is taking place, where TDP interventions are operating.

212. Based on the qualitative matrix above, fieldwork is proposed with the following groups, for which specific tools will be developed following the completion of the quantitative surveys, in order to reflect emergent findings:

- Children (informed consent required), headteachers, TDP-trained teachers and other teachers in TDP in-service schools, interviews using a case study approach (i.e. following up the same individuals unless they are transferred). Focus groups may be used as appropriate and decided by researchers, but are not currently planned.
- Observations of lessons,
- Observation of teacher training sessions,
- LGEA, SUBEB, ERDC, teacher facilitators and teacher development team members, and ministry, UBEC and NERDC staff, interviews
- TDP staff, interviews
- ESSPIN, GEP3 staff, interviews
- DFID staff, interviews.

213. The following table lists each respondent, the areas of discussion, by year. Data collection is expected to be interviews and observations of training sessions and lessons.

Table 12. Qualitative question areas, by respondent and year**2015****Attendance register**

Perceptions of regular attendance by students and teachers; % days attended

Classroom observations

Number of teachers demonstrating adequate classroom materials

Perceptions of adequacy of classroom materials

Perceptions of peer support

Perceptions of student ability to learn in language of instruction

Perceptions of students understanding materials and curriculum

Students reporting understanding materials and curriculum

Students understanding materials and curriculum

Teacher and student perceptions of adequacy of materials

Teachers being able to use TDP technology

Teachers using English to teach P4-6

Teachers using Hausa to teach maths 1-3

EDRC

EDRC is the appropriate institution for material and technology

ESSPIN staff

Is TDP designed appropriately in relation to GEP3 and ESSPIN?

FMOE

Are FMOE, UBEC and NERDC appropriate custodians of in-service teacher training in Nigeria?

Are SUBEBs appropriate institutional homes for in-service training in each state?

Do state governments/SUBEBs consider improving in-service teacher training through the TDP model a priority?

Has TDP adequately designed an approach to account for variations in states' capacity to adopt the TDP?

Is teacher effectiveness reform a policy priority in Nigeria?

FMOE, UBEC, NERDC

Is a programme of continuous support to teachers realistic and aligned with the policy context in Nigeria?

GEP3 staff

Is TDP designed appropriately in relation to GEP3 and ESSPIN?

Head teachers

% teachers and head teachers receiving TDP training; frequency of attendance

% teachers with Smartphone; % teachers with electricity at home; % teachers reporting using 'trainer-in-pocket'; % schools with facilities for technology

Head teacher perceptions of incentives/motivation

LGEA

Teacher educator and teacher facilitator perception of adequate time and support

Teachers and government officials reporting TDP materials appropriate, for male and female teachers

NERDC

Are FMOE, UBEC and NERDC appropriate custodians of in-service teacher training in Nigeria?

Are SUBEBs appropriate institutional homes for in-service training in each state?

Do state governments/SUBEBs consider improving in-service teacher training through the TDP model a priority?

Has TDP adequately designed an approach to account for variations in states' capacity to adopt the TDP?

Is teacher effectiveness reform a policy priority in Nigeria?

School records

STR; perceptions of adequacy of class size

SUBEB

Are SUBEBs appropriate institutional homes for in-service training in each state?

Do state governments/SUBEBs consider improving in-service teacher training through the TDP model a priority?

EDRC is the appropriate institution for material and technology

Exogenous changes

Has TDP adequately designed an approach to account for variations in states' capacity to adopt the TDP?

Is teacher effectiveness reform a policy priority in Nigeria?

Is the application of the TDP model across TDP states appropriate and consistent with the state policy context?

Teacher educator and teacher facilitator perception of adequate time and support

Teachers and government officials reporting TDP materials appropriate, for male and female teachers

TDP staff

Efficiency of TDP's organisational management and set up

TDP taught students

Perceptions of adequacy of classroom materials

Perceptions of adequate support at home

Perceptions of peer support

Perceptions of regular attendance by students and teachers; % days attended

Perceptions of student ability to learn in language of instruction

Perceptions of students understanding materials and curriculum

STR; perceptions of adequacy of class size

Students reporting understanding materials and curriculum

Students understanding materials and curriculum

Teacher and student perceptions of adequacy of materials

TDP trained teachers

% teachers and head teachers receiving TDP training; frequency of attendance

Head teacher perceptions of incentives/motivation

Number of teachers demonstrating adequate classroom materials

Perceptions of adequacy of classroom materials

Perceptions of adequate support at home

Perceptions of peer support

Perceptions of regular attendance by students and teachers; % days attended

Perceptions of student ability to learn in language of instruction

Perceptions of students understanding materials and curriculum

STR; perceptions of adequacy of class size

Students understanding materials and curriculum

Teacher and student perceptions of adequacy of materials

Teacher educator and teacher facilitator perception of capacity to absorb training

Teacher facilitator, head teacher and teacher perceptions of adequacy of training

Teachers and government officials reporting TDP materials appropriate, for male and female teachers

Teachers being able to use TDP technology

Teachers have the basic language, subject and pedagogical skills to absorb the new knowledge and skills available from TDP;

Teachers using English to teach P4-6

Teachers using Hausa to teach maths 1-3

Teacher facilitators

Teacher educator and teacher facilitator perception of adequate time and support

Teacher educator and teacher facilitator perception of capacity to absorb training
Teacher facilitator, head teacher and teacher perceptions of adequacy of training

Training observations

Teacher educator and teacher facilitator perception of capacity to absorb training
Teacher facilitator, head teacher and teacher perceptions of adequacy of training

UBEC

Are FMoE, UBEC and NERDC appropriate custodians of in-service teacher training in Nigeria?

Are SUBEBs appropriate institutional homes for in-service training in each state?

Do state governments/SUBEBs consider improving in-service teacher training through the TDP model a priority?

Has TDP adequately designed an approach to account for variations in states' capacity to adopt the TDP?

Is teacher effectiveness reform a policy priority in Nigeria?

2016

Attendance register

Perceptions of regular attendance by students and teachers; % days attended

Classroom observations

Perception of usefulness of TDP materials

Perceptions of adequacy of classroom materials

Perceptions of other teacher effectiveness; self-reported application of model

Perceptions of peer support

Perceptions of student ability to learn in language of instruction

Perceptions of students understanding materials and curriculum

Teacher and student perceptions of adequacy of materials

Teachers reporting/displaying adequate intrinsic motivation

DFID staff

Perception of ESSPIN and GEP3 capacity to apply model

ESSPIN staff

Perception of ESSPIN and GEP3 capacity to apply model

GEP3 staff

Perception of ESSPIN and GEP3 capacity to apply model

Head teachers

% teachers and head teachers receiving TDP training; frequency of attendance

% teachers in TDP schools transferred in previous year

% teachers with appropriate qualifications; % teachers receiving training in last three years

Head teacher perceptions of incentives/motivation

Number of head teacher meetings with other teachers; number of head teacher lesson observations of other teachers; number of head teachers taking action to reduce absenteeism; perceptions of wider application by head teachers

Number of head teacher meetings with teachers; number of head teacher lesson observations; number of head teachers taking action to reduce absenteeism

Perceptions of other teacher capacity

Perceptions of other teacher effectiveness; self-reported application of model

Reported sharing of information by head teachers

LGEA

Teacher educator and teacher facilitator perception of adequate time and support

Teachers and government officials reporting TDP materials appropriate, for male and female teachers

Other students

Perceptions of other teacher capacity

Perceptions of other teacher effectiveness; self-reported application of model

Other teachers

Number of head teacher meetings with other teachers; number of head teacher lesson observations of other teachers; number of head teachers taking action to reduce absenteeism; perceptions of wider application by head teachers

Perceptions of other teacher capacity

Perceptions of other teacher effectiveness; self-reported application of model

Reported sharing of information by head teachers

Other teachers classroom observations

Perceptions of other teacher capacity

School records

STR; perceptions of adequacy of class size

SUBEB

Exogenous changes

Teacher educator and teacher facilitator perception of adequate time and support

Teachers and government officials reporting TDP materials appropriate, for male and female teachers

TDP staff

Efficiency of TDP's organisational management and set up

Perception of ESSPIN and GEP3 capacity to apply model

TDP taught students

Perception of usefulness of TDP materials

Perceptions of adequacy of classroom materials

Perceptions of adequate support at home

Perceptions of peer support

Perceptions of regular attendance by students and teachers; % days attended

Perceptions of student ability to learn in language of instruction

Perceptions of students understanding materials and curriculum

STR; perceptions of adequacy of class size

Teacher and student perceptions of adequacy of materials

TDP trained teachers

% teachers and head teachers receiving TDP training; frequency of attendance

% teachers with appropriate qualifications; % teachers receiving training in last three years

Head teacher perceptions of incentives/motivation

Number of head teacher meetings with teachers; number of head teacher lesson observations; number of head teachers taking action to reduce absenteeism

Perception of usefulness of TDP materials

Perceptions of adequacy of classroom materials

Perceptions of adequate support at home

Perceptions of peer support

Perceptions of regular attendance by students and teachers; % days attended

Perceptions of student ability to learn in language of instruction

Perceptions of students understanding materials and curriculum

STR; perceptions of adequacy of class size

Teacher and student perceptions of adequacy of materials

Teacher educator and teacher facilitator perception of capacity to absorb training

Teacher facilitator, head teacher and teacher perceptions of adequacy of training

Teachers and government officials reporting TDP materials appropriate, for male and female teachers

Teachers have the basic language, subject and pedagogical skills to absorb the new knowledge and skills available from TDP;

Teachers reporting adequate extrinsic motivation

Teachers reporting/displaying adequate intrinsic motivation

Teacher facilitators

Teacher educator and teacher facilitator perception of adequate time and support

Teacher educator and teacher facilitator perception of capacity to absorb training

Teacher facilitator, head teacher and teacher perceptions of adequacy of training

Training observations

Teacher educator and teacher facilitator perception of capacity to absorb training

Teacher facilitator, head teacher and teacher perceptions of adequacy of training

2017**Attendance register**

Perceptions of regular attendance by students and teachers; % days attended

Classroom observations

Perception of usefulness of TDP materials

Perceptions of adequacy of classroom materials

Perceptions of other teacher effectiveness; self-reported application of model

Perceptions of peer support

Perceptions of student ability to learn in language of instruction

Perceptions of students understanding materials and curriculum

Teacher and student perceptions of adequacy of materials

DFID staff

Perception of appropriateness of model

Perception of institutional capacity to apply TDP model

EDRC

EDRC is the appropriate institution for material and technology

FMOE

Are SUBEBs appropriate institutional homes for in-service training in each state?

Do state governments/SUBEBs consider improving in-service teacher training through the TDP model a priority?

FMOE, UBEC and NERDC are appropriate custodians of in-service teacher training in Nigeria

In-service teacher training reform is a priority for teacher effectiveness reform in Nigeria

Perception of appropriateness of model

Perception of institutional capacity to apply TDP model

Perception of SUBEB capacity to apply TDP model

Perceptions of application of model in other states

Perceptions of maintenance of TDP approach and practice

Perceptions of SUBEB and LGEA capacity to maintain TDP approach and practice

Teacher effectiveness reform is a policy priority in Nigeria

Head teachers

% teachers and head teachers receiving TDP training; frequency of attendance

% teachers in TDP schools transferred in previous year

% teachers with appropriate qualifications; % teachers receiving training in last three years

Head teacher perceptions of incentives/motivation

Number of head teacher meetings with other teachers; number of head teacher lesson observations of other teachers; number of head teachers taking action to reduce absenteeism; perceptions of wider application by head teachers

Number of head teacher meetings with teachers; number of head teacher lesson observations; number of head teachers taking action to reduce absenteeism

Perception of longevity of technology

Perceptions of head teacher and teacher interest and capacity maintain TDP approach and practice

Perceptions of maintenance of TDP approach and practice

Perceptions of other teacher capacity

Perceptions of other teacher effectiveness; self-reported application of model
Reported sharing of information by head teachers

LGEA

Perceptions of maintenance of TDP approach and practice
Perceptions of SUBEB and LGEA capacity to maintain TDP approach and practice
Teacher educator and teacher facilitator perception of adequate time and support
Teachers and government officials reporting TDP materials appropriate, for male and female teachers

NERDC

Are SUBEBs appropriate institutional homes for in-service training in each state?
Do state governments/SUBEBs consider improving in-service teacher training through the TDP model a priority?
FMOE, UBEC and NERDC are appropriate custodians of in-service teacher training in Nigeria
In-service teacher training reform is a priority for teacher effectiveness reform in Nigeria
Perception of appropriateness of model
Perception of institutional capacity to apply TDP model
Perception of SUBEB capacity to apply TDP model
Teacher effectiveness reform is a policy priority in Nigeria

Other students

Perceptions of other teacher capacity
Perceptions of other teacher effectiveness; self-reported application of model

Other teachers

Number of head teacher meetings with other teachers; number of head teacher lesson observations of other teachers; number of head teachers taking action to reduce absenteeism; perceptions of wider application by head teachers
Perceptions of other teacher capacity
Perceptions of other teacher effectiveness; self-reported application of model
Reported sharing of information by head teachers

Other teachers classroom observations

Perceptions of other teacher capacity

School records

STR; perceptions of adequacy of class size

SUBEB

Are SUBEBs appropriate institutional homes for in-service training in each state?
Do state governments/SUBEBs consider improving in-service teacher training through the TDP model a priority?
EDRC is the appropriate institution for material and technology
Exogenous changes
In-service teacher training reform is a priority for teacher effectiveness reform in Nigeria
Perception of appropriateness of model
Perception of institutional capacity to apply TDP model
Perception of SUBEB capacity to apply TDP model
Perceptions of application of model in other schools in TDP states
Perceptions of maintenance of TDP approach and practice
Perceptions of SUBEB and LGEA capacity to maintain TDP approach and practice
Teacher educator and teacher facilitator perception of adequate time and support
Teacher effectiveness reform is a policy priority in Nigeria
Teachers and government officials reporting TDP materials appropriate, for male and female teachers

TDP staff

Efficiency of TDP's organisational management and set up
Flexibility of training materials

Perception of appropriateness of model

Perception of institutional capacity to apply TDP model

TDP taught students

Perception of usefulness of TDP materials

Perceptions of adequacy of classroom materials

Perceptions of adequate support at home

Perceptions of peer support

Perceptions of regular attendance by students and teachers; % days attended

Perceptions of student ability to learn in language of instruction

Perceptions of students understanding materials and curriculum

STR; perceptions of adequacy of class size

Teacher and student perceptions of adequacy of materials

TDP trained teachers

% teachers and head teachers receiving TDP training; frequency of attendance

% teachers with appropriate qualifications; % teachers receiving training in last three years

Head teacher perceptions of incentives/motivation

Number of head teacher meetings with teachers; number of head teacher lesson observations; number of head teachers taking action to reduce absenteeism

Perception of longevity of technology

Perception of usefulness of TDP materials

Perceptions of adequacy of classroom materials

Perceptions of adequate support at home

Perceptions of head teacher and teacher interest and capacity maintain TDP approach and practice

Perceptions of maintenance of TDP approach and practice

Perceptions of peer support

Perceptions of regular attendance by students and teachers; % days attended

Perceptions of student ability to learn in language of instruction

Perceptions of students understanding materials and curriculum

STR; perceptions of adequacy of class size

Teacher and student perceptions of adequacy of materials

Teacher educator and teacher facilitator perception of capacity to absorb training

Teacher facilitator, head teacher and teacher perceptions of adequacy of training

Teachers and government officials reporting TDP materials appropriate, for male and female teachers

Teachers have the basic language, subject and pedagogical skills to absorb the new knowledge and skills available from TDP;

Teachers reporting adequate extrinsic motivation

(blank)

Teacher facilitators

Perceptions of maintenance of TDP approach and practice

Teacher educator and teacher facilitator perception of adequate time and support

Teacher educator and teacher facilitator perception of capacity to absorb training

Teacher facilitator, head teacher and teacher perceptions of adequacy of training

Training observations

Teacher educator and teacher facilitator perception of capacity to absorb training

Teacher facilitator, head teacher and teacher perceptions of adequacy of training

UBEC

Are SUBEBs appropriate institutional homes for in-service training in each state?

Do state governments/SUBEBs consider improving in-service teacher training through the TDP model a priority?

FMoE, UBEC and NERDC are appropriate custodians of in-service teacher training in Nigeria

In-service teacher training reform is a priority for teacher effectiveness reform in Nigeria
 Perception of appropriateness of model
 Perception of institutional capacity to apply TDP model
 Perception of SUBEB capacity to apply TDP model
 Perceptions of application of model in other states
 Perceptions of maintenance of TDP approach and practice
 Perceptions of SUBEB and LGEA capacity to maintain TDP approach and practice
 Teacher effectiveness reform is a policy priority in Nigeria

214. Sampling for the qualitative fieldwork can be divided into three approaches. Most simply, for the TDP and DFID staff, the fieldwork will interview everyone who is relevant (actual saturation). Second, for the LGEA (from in-service treatment cluster LGAs), SUBEB, EDRC and ministry, UBEC and NERDC staff, for whom it is not always clear a priori who is responsible for or informed about different issues, the fieldwork will take a snowball approach, first working with designated staff with which the TDP has connections, and then asking them for connections with other staff that they think will know or care about issues of teaching and teacher effectiveness, and so on until theoretical saturation is reached. These lists will be retained for subsequent rounds of fieldwork, but the snowballing refreshed.

215. Finally, the sampling of schools, head teachers, teachers, teacher facilitators and children for in-service evaluation is most complex. Qualitative fieldwork will take place in schools in the in-service quantitative survey treatment clusters. Case study schools will be selected on the basis of the baseline MLA/TDNA survey, identifying schools with comparatively good and poor levels of learning where possible (the sample is not designed to differentiate learning or teaching at the level of a school, as standard errors around school measurements are likely to be too large to differentiate, so this will be merely a least bad basis for guessing quality). Four schools will be selected in each state (two performing well, two poorly).

216. In these case study schools, a panel of students and TDP teachers will be identified. Four students will be starting p1, and four others starting p3, in order to track changes in learning over three years for two different cohorts using different TDP materials. Three teachers will initially be those teaching p1-3 and supported by TDP (matching the quantitative evaluation), but other teachers that do not receive TDP support will be added in subsequent years. Headteachers will also be interviewed. These panels will be maintained with annual visits, interviews and classroom observations for teachers. One non-TDP trained teacher will be identified at random from the teachers' list for interviews and classroom observations. If students or teachers leave, they will be replaced. In each school, the teacher facilitators and educators that work with that school, and the local LGEA, will be interviewed and training sessions observed. This is set out in the table below.

Table 13. Qualitative sampling

Qualitative respondent	Number per state per round	Total per round
Attendance register	4	12
Classroom observations	16	48
DFID staff	1+	1+
EDRC	1+	1+
ESSPIN staff	1+	1+
FMOE	1+	1+
GEP3 staff	1+	1+
Head teachers	4	12
LGEA	1	3
NERDC	1+	1+
Other students	16	48
Other teachers	4	12
Other teachers classroom observations	4	12
School records	4	12
SUBEB	1+	1+
TDP staff	1+	1+
TDP taught students	12	36
TDP trained teachers	12	36
Teacher facilitators	4	12
Training observations	4	12
UBEC	1+	1+
Grand Total	85	255

217. In the event that insecurity, the elections or Ebola significantly disrupts the TDP or the evaluation, additional qualitative fieldwork may be planned and conducted as a replacement for the 2015 midline survey. This is discussed briefly in section 5.4.

4.2.3 Limitations

218. Limitations of the rigorous impact evaluation have been set out above. The main limitations of the qualitative part of evaluation are presented below. Column one states the possible limitations, whilst column two explains why this is limiting and outlines how the impact evaluation will seek to mitigate this.

Table 14. Possible limitations of qualitative component

Possible limitation	Why this is limiting and what can be done
Inference beyond the selected research sites is limited.	The case study approach may suffer from having selected schools that are atypical. This may mean that findings will not be applicable to other TDP areas or beyond. This is mitigated by combining the qualitative information with quantitative surveys in a wider selection of schools.
Attrition of the qualitative panels	This would reduce the explanatory power of the qualitative work in schools as relationships of trust must be rebuilt and there is a shorter timeframe to interact with research participants. Those who leave the panel will be replaced.
The TDP design may change in phase 2	The formative part of the evaluation may lead to changes in the way in which the TDP is designed and implemented in phase 2. This would reduce the validity of the qualitative (and quantitative findings) from phase 1. If this occurs, it may be worth considering some additional funding for qualitative case studies in phase 2 states.

4.2.4 Process evaluation

219. The process evaluation questions (see questions Effi6-Effi23 in the evaluation framework in Annex G) will be answered using rigorous factual analysis of principally qualitative data and programme monitoring data. The process evaluation will test whether the TDP is achieving its outputs as expected, and this underpins and explains the analysis of the achievement of outcomes and impact.

220. The first set of process evaluation questions use programme monitoring data to assess whether the programme's results were achieved on time and in full. (These questions are also checked in the TDP Annual Reviews). The second set of process evaluation questions explore how the TDP's organisational set up and management facilitated programme delivery, and these will draw additionally from qualitative interviews with TDP staff and other key stakeholders. These interviews will be triangulated against programme data and documents. The process evaluation is an important part of the impact evaluation – because the achievement of output is critical to the achievement of outcomes and impact – but the focus of this evaluation is not testing operational procedures but assessing their impact. The process evaluation is therefore a comparatively small part of the evaluation overall.

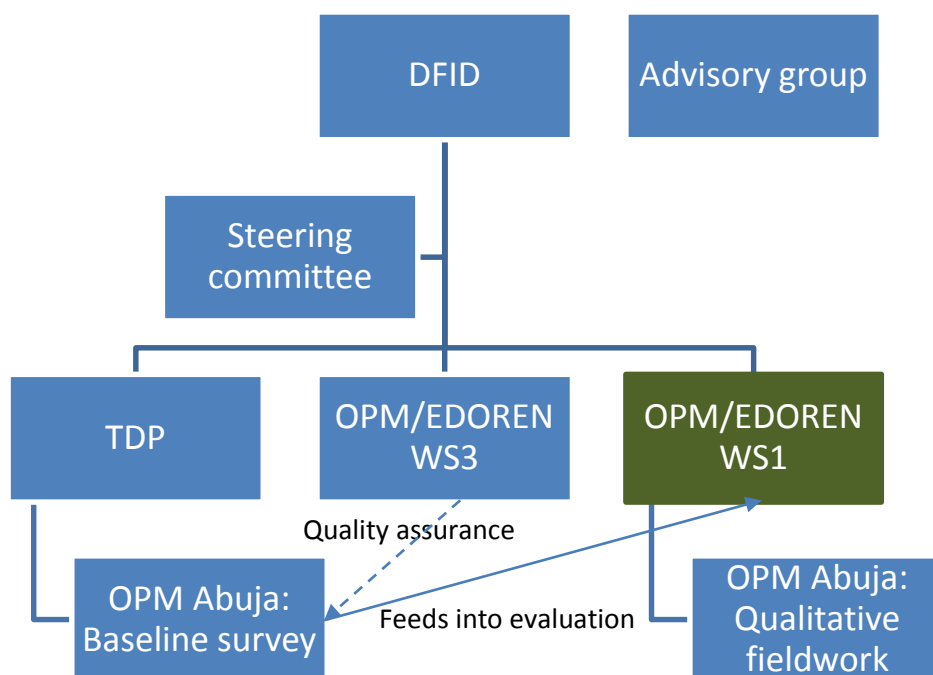
5 Evaluation governance, management, independence and ethics

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. The evaluation will be governed by a steering group composed of DFID staff, a TDP representative, and an EDOREN representative. The independence of the evaluation is assured through transparency and rigorous peer review.

5.1 Governance and management

221. This evaluation framework and the evaluation that it sets out are managed and implemented by EDOREN, a DFID-funded project managed by Oxford Policy Management (OPM), and specifically EDOREN workstream 1. EDOREN workstream 1 is responsible for the design of the evaluation framework and the implementation of the evaluation. The evaluation will draw extensively on information and data gathered by OPM's Abuja office under contract directly from the TDP (for the baseline survey), as well as data collected by OPM Abuja under the auspices of our EDOREN contract (for qualitative work and follow up surveys). The OPM Abuja baseline survey group will report regularly to the EDOREN workstream 1 TDP evaluation team to ensure that the baseline meets the requirements of the evaluation framework as well as of the TDP. In addition, EDOREN's workstream 3 on statistical quality will independently review the quality of quantitative data collections conducted by OPM Abuja.

Figure 6: TDP evaluation governance arrangements



222. The evaluation will be governed by DFID. Specifically, this entails that the evaluation 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.

223. In addition, it is recommended that an evaluation steering committee be formalised, composed of DFID Nigeria education and results advisers, TDP and EDOREN staff – a group that already meets regularly. This committee could formally meet every six months to discuss evaluation progress, and would be responsible for peer review and quality assurance (in addition to EDOREN's internal quality assurance

processes). The designated EDOREN workstream 1 lead for the TDP evaluation would report to this committee every six months.

224. An advisory group of secondary users should also 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 should meet annually to discuss evaluation progress as reported by the EDOREN workstream 1 lead for the TDP evaluation, and would offer advice to next steps.

5.2 Independence, credibility and usefulness

225. This section briefly discusses how the evaluation achieves the principles of independence, credibility and usefulness, and appropriate participation. These principles are detailed in Annex D.

226. 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 (such as in this evaluation framework and in Bennell 2013) on the nature of its interventions. EDOREN is also both designing this evaluation framework and 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's evaluation. Our approach to mitigating these concerns is to ensure that 1) ensure all evaluation outputs are peer reviewed, 2) data and analytical approaches for both quantitative and qualitative data are transparent (i.e. publicly available for scrutiny), and 3) to contract independent annual review teams as per EDOREN's inception report.

227. 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.

228. 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 midline reports in 2015 and 2016 will provide important evidence for adjustments to TDP phase 2. Our approach to ensuring the relevance of the evaluation questions and proposed priorities is to consult widely around this document. Credibility and reliability will be achieved through the steps outlined above, and through the use of high quality expertise in conducting the evaluation.

229. In addition, TDP's evaluation will aim to fulfil the principles of usefulness and participation 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 approach, and the evaluation's proposed approach to stakeholder engagement, for which see section 3.7.

230. This draft of the evaluation framework is intended for consumption by TDP and DFID Nigeria to validate the overall approach and the detail around the interventions.

231. In addition, ethical evaluation principles are important, and run through the evaluation design below. They will be explored in more detail in the documentation around individual data collection activities. One area of particular importance for this evaluation will be ensuring that children are asked age-appropriate questions and appropriate consent is sought, and this will follow UNICEF guidelines on conducting research with children. In addition, research proposals will be passed by OPM's ethical review board and the relevant review boards in Nigeria. These principles are operationalised below.

5.3 Research ethics

232. The specific operationalization of research ethics principles set out in Annex section D.2.2 will be presented in more detail in the documentation around individual data collection activities. One area of particular importance for this evaluation will be ensuring that children are asked age-appropriate questions and appropriate consent is sought, and this will follow UNICEF guidelines on conducting research with children. In addition, research proposals will be passed by OPM's ethical review board and the relevant review boards in Nigeria, probably the National Health Research Ethics Committee.

233. All TDP fieldwork will include:

- **Informed consent from participants.** For children, permission will be sought from both the primary caregiver (a teacher if the child is in school) as well children themselves.
- **Anonymity:** All datasets will be anonymised, in the sense that all names of respondents are removed before the data is shared publically.
- **Ensuring the safety of participants:** Fieldworkers will be local to areas in which they are assigned. In addition fieldwork supervisors will support the fieldwork manager in monitoring local security concerns.

234. In addition, no work with children will be carried out unless it both meets the ethical principles set out in Annex section D.2.2, and is absolutely necessary for the evaluation.

5.4 Risks

235. There are three major risks to implementation of the TDP which would affect the evaluation substantially:

- The TDP operates in states that have generalised insecurity. Some schools with which the TDP planned to work are inaccessible due to insecurity, and the replacement of these schools would have impacts for the evaluation, as well as the representativeness of the findings. Attacks directly targeted on schools would significantly and directly affect the TDP's in-service work. The bombings of Colleges of Education (such as recently in Niger State) have a substantial impact on the TDP's pre-service programme, but also on the in-service programme as some members of the Teacher Development Teams are lecturers in Colleges of Education.
- The elections in February 2015 could lead to school closure. In August and September 2014, the Federal Ministry of Education announced the closure of all schools in response to the outbreak of Ebola in Lagos and Port Harcourt. This might have had substantial impacts on the programme timeline and the evaluation baseline. However, the schools re-opened sufficiently soon to allow both the programme and evaluation fieldwork to continue without substantial delay. Currently, it is expected that the elections will not cause substantial problems, but the risk is there, and the evaluation does not expect to conduct any fieldwork during the election period.
- A further outbreak of Ebola in northern Nigeria could have catastrophic consequences for the programme, and therefore the evaluation. At present, Nigeria has been declared Ebola-free by the World Health Organisation. However, there remains a risk that the disease re-enters the country. The risk to the north is probably lower than to Lagos or other major cities, given comparatively low population density and international travel, but remains extant.

236. The evaluation's approach to managing these risks is to carefully monitor them through EDOREN's Abuja office and the TDP's offices, and if these issues appear, to discuss with DFID, the TDP and the evaluation steering committee. If a delay to the programme takes place, it may be sensible to delay fieldwork, depending on the duration and scale of the impact on programme operations. If the TDP is

forced to make major changes or fieldwork become impossible, this would require more substantial change to the evaluation, possibly dropping rounds of fieldwork or putting more emphasis on the qualitative work. At present, however, the likelihood of these risks materialising is judged as low.

237. If this judgement turns out to be over-optimistic and the risks around insecurity, the elections or Ebola significantly disrupt the TDP's activities of the evaluation, it may be necessary to drop the 2015 midline surveys. In this case, Phase 2 activities would be informed by additional qualitative fieldwork. The details of this additional fieldwork would be discussed with the evaluation steering committee with a decision made by July 2015.

6 Strategy for stakeholder engagement and communication

In this chapter we discuss the strategy for engaging stakeholders throughout the various stages of the evaluation process, and plans for disseminating 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, development of tools and data collection. The first and second sections of this chapter outline the roles of stakeholders, and their involvement in the development of the evaluation framework. In the final section we discuss the plan for communicating evaluation results and key considerations.

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

6.1 Role and involvement of stakeholders

239. 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, the conduct of the evaluation, and communication of evaluation findings. Involving stakeholders at each stage of the evaluation will help to ensure that (1) there is a common understanding of the problem being addressed by the intervention; (2) the right questions are asked; (3) the questions are appropriately phrased; and (4) 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.

240. The main purpose of this evaluation is learning, i.e. learning what works to support teachers in resource-poor contexts. Therefore, the principal stakeholders for the evaluation are:

- I. DFID Nigeria,
- II. TDP,
- III. the State governments of the Katsina, Jigawa and Zamfara – especially State Ministries of Education and State Universal Basic Education Board (SUBEBs);
- IV. NCCE and Colleges of Education;
- V. Other education projects especially GEP 3 and ESSPIN;
- VI. organisations (primarily DFID) seeking to improve teacher competence and subsequently learning elsewhere in the world,
- VII. international researchers on education, and
- VIII. Nigerian education policymakers and researchers.

241. Involving key stakeholders in the design of the evaluation framework and the conduct of the evaluation provides perspective to ensure a credible, high quality and useful evaluation; and contributes to program logic and formulation of key evaluation questions.

242. 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 ensures 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 evaluations findings.

243. Key stakeholders will be involved in the conduct of the evaluation in the following ways:

- i. Interactive discussions with TDP to finalise data collection tools and methodologies to ensure all necessary questions are asked and there is consensus on every aspect of the approach

- iii. Rapid feedback of results during data collection shall be provided to the local stakeholders including schools, teachers, students and communities. The aim of this would be to validate the message coming out from the field and improve the quality of research, as well as communicate to respondents that their opinions were understood (OPM, 2013).
- iv. Interactive problem discussion and analysis with TDP, and representatives of state government institutions to discuss issues identified during the evaluation and identify possible solutions.
- v. Feedback of results to TDP, DFID Nigeria, and the Governments of Katsina, Jigawa and Zamfara to ensure they are taken into account for other (current or future) teacher development programmes aimed at improving teacher competency and broader learning outcomes.
- vi. Feedback of results to DFID Nigeria; international organisations (primarily DFID) seeking to improve teacher competencies elsewhere in the world; and international education policymakers and researchers to help make more informed decisions about whether to apply the TDP model of teacher development in other contexts.

244. Stakeholders will be involved in two broad phases: development of the evaluation framework and communication of evaluation findings.

245. 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.

6.2 Stakeholder involvement in framework development

246. Stakeholders will be involved in the development of the evaluation framework in various ways:

- I. DFID Nigeria. Education advisers and the results adviser will have final signoff on the evaluation framework.
- II. TDP Results and Evidence Lead Bukola Oyinloye was consulted on the development of the framework
- III. International researchers within OPM and EDORENs networks.

247. Consultative discussions were held with the DFID team to understand the main aims of the intervention and the key activities within each component and main uses of an evaluation. Discussions were also had with DFID to clarify the main uses and purposes of the evaluation.

248. In the next stage, a first draft of this report will be shared with the TDP team, revised and shared with DFID, and then subsequently with a wider group of external peer reviewers and stakeholders. This may include a discussion in Abuja on a draft of the report.

6.3 Strategies to communicate findings

249. 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 communication of evaluation findings will be developed in line with TDPs communication strategy and EDORENs overall dissemination plan.

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

- I. Ensure 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.
- II. Evidence is 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 key findings in non-technical language, supported by technical annexes (Better evaluation, 2013).
- III. Ensure 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.
- IV. Appropriate strategies of stakeholder engagement are applied to promote engagement with the evidence and subsequent use: Meetings (whether of stakeholder, one-to-one, or teams) are a key mechanism for sharing knowledge and learning, and facilitating stakeholder understanding of the evidence and how this might affect their practise 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 on education related issues.

References

- Adekola, O.A., 2007a. Language, literacy and learning in primary schools: implications for teacher development programs in Nigeria. World Bank Publications.
- Adekola, O.A., 2007b. Language, Literacy, and Learning in Primary Schools: Implications for Teacher Development Programs in Nigeria, World Bank Working Papers. The World Bank.
- Adelabu, M.A., 2005. Teacher motivation and incentives in Nigeria.
- Allsop, T., Howard, L., 2009. ESSPIN an Assessment of Teacher Education in Jigawa State. (No. JG 302). ESSPIN, Abuja, Nigeria.
- Aslam, M., Kingdon, G., 2011. What can teachers do to raise pupil achievement? Econ. Educ. Rev. 30, 559–574.
- Boulton, D., Jackson, P., Oliver, D., 2009. ESSPIN School Case Study Reports – Kaduna, Kano & Kwara States (No. ESSPIN 012).
- Cartwright and Hardie (2012), Evidence-based policy.
- Chetty, R., Friedman, J.N., Hilger, N., Saez, E., Schanzenbach, D.W., Yagan, D., 2011. How does your kindergarten classroom affect your earnings? Evidence from Project STAR. Q. J. Econ. 126, 1593–1660.
- DAC, (2010), Evaluation quality standards
- DFID, 2013. Education position paper - improving learning, expanding opportunities (Position paper). DFID, London.
- DFID Nigeria, 2012. Teacher Development Programme (TDP) Business Case.
- Federal Ministry of Education, Nigeria, 2004. National Policy on Education, 4th Ed.
- Johnson, D., and Hsieh, J., (2014), ‘A diagnostic assessment of the professional development needs of teachers in Nigeria: the characteristics, knowledge and skills of teachers in Katsina State’, report number 4 for GEP3. UNICEF.
- Hanushek, E.A., Rivkin, S.G., 2006. Chapter 18 Teacher Quality, in: Handbook of the Economics of Education. Elsevier, pp. 1051–1078.
- Humphreys, S., Crawford, L., 2014. Review of the literature on basic education in Nigeria.
- Igwe, R.O., Rufai, S.A., 2012. An Evaluation of Teachers’ Service Delivery in Nigeria: Measures for Programme Accountability and Improvement. Humanity Soc. Sci. J. 7, 121–127.
- Jigawa State, Ministry of Education, Science and Technology, 2012. JIGAWA STATE: ANNUAL SCHOOL CENSUS REPORT 2011/2012. Jigawa SUBEB & MOEST EDUCATION MANAGEMENT INFORMATION SYSTEM UNIT (EMIS).
- Johnson, D., Hsieh, J., 2014. An Assessment of the Development Needs of Teachers in Nigeria - Katsina State Case Study.
- Nannyonjo, H., 2007. Education Inputs in Uganda: An Analysis of Factors Influencing Learning Achievement in Grade Six, World Bank Working Papers. The World Bank.
- Orr D, Westbrook J, Pryor J, Durrani N, Sebba J, Adu-Yeboah C (2013) *What are the impacts and cost effectiveness of strategies to improve performance of untrained and under-trained teachers in the classroom in developing countries?* London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.
- Sherry, H, 2008. Teachers’ voice: a policy research report on teachers’ motivation and perceptions of their profession in Nigeria. VSO, London, UK.
- TDP, 2014a. TDP Inception report.

TDP, 2014b. TDP In-service teacher training strategy.

TDP, 2014c. TDP Pre-service implementation strategy.

TDP, 2014d. TDP Results and Evidence strategy.

Thomas, H.G., 2011. ESSPIN Study of Teacher Management and Deployment (Assignment Report No. ESSPIN 328).

UNESCO, 2014. Teaching and learning - achieving quality for all (EFA Global Monitoring Report No. 201 3/4). UNESCO.

Vogel, I., (2012), 'Review of the use of theory of change in international development', Review paper for the Department of International Development. DFID.

Westbrook, J., 2013. Pedagogy 2013 Westbrook report.pdf.

White, H., (2009), 'Theory-based impact evaluation: principles and practice', 3ie Working Paper 3. International Initiative on Impact Evaluation.

Yoshikawa, H., McCartney, K., Myers, R.G., Bub, K.L., Lugo-Gil, J., Ramos, M.A., Knaul, F., Centre, U.I.R., 2007. Early Childhood Education in Mexico: Expansion, quality improvement, and curricular reform (Innocenti Working Paper No. inwopa07/41). UNICEF Innocenti Research Centre.

Annex A Terms of reference

WORKSTREAM 1

TEACHER DEVELOPMENT PROGRAMME EVALUATION FRAMEWORK

Terms of Reference

EDOREN

12th May 2014

A.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.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 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. Following this, it is expected (but not yet contracted) that EDOREN will conduct the evaluation of TDP. EDOREN will continue to advise TDP on their M&E strategy.

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

A.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 DEEPEN'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.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.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 staff before proposing them. Following discussions with the TDP team and using the TDP M&E strategy, it seems that the in-service component could be evaluated using a quasi-experimental design, and the pre-service component would be evaluated using research approaches with smaller 'N's. Both these approaches would need to be validated.

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 TDP to develop an initial sample of 48 schools for a baseline for the in-service evaluation. His sampling note (due w/c 13th May), 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. 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 are likely to 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 six. In some schools there are up to 50 primary teachers. It is expected that teachers will be selected by headteachers 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 Monitoring and Evaluation strategy. This currently proposes a series of surveys at school and household level (DEEPEN 2014). EDOREN proposed to link these surveys (Bano and Bennell 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 DEEPEN 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. 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 is it planned to evaluate the intervention?
4. What are the key evaluation questions, organised by DAC criteria and drawing on the logical framework and theory of change, and the initial review of evaluability conducted by EDOREN?
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 is the role 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), 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.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	30 th June 2014
Draft evaluation framework including 2 page executive summary and full indicator framework annex	One week after receiving DFID comment (expected final review submission date – 23 rd July 2014)

A.7 Timeframe

The framework is to be completed by 30th June 2014. The time allocations are as follows:

- a. Team Leader – Evaluation Expert – 20 days
- b. Teacher Education Expert – 15 days
- c. Analysts/Nigeria education experts - 30 days

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

A.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. The team leader will work closely with Ian MacAuslan, EDOREN's Review and Evaluation leader, throughout the process. 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.9 Coordination and Logistics

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

Coordination with TDP 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 the TDP team.

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.10 Reference Documents

Bennell (2013), 'Draft comments on TDP Design', EDOREN.

www.BetterEvaluation.org

DAC, (2010), Evaluation quality standards

DFID evaluation policy 2013 (<https://www.gov.uk/government/publications/dfid-evaluation-policy-2013>), particularly for DFID evaluation principles

DFID ethics principles for research and evaluation (<https://www.gov.uk/government/publications/dfid-ethics-principles-for-research-and-evaluation>)

DFID (no date), 'Evaluation Handbook 5', document made available to EDOREN.

DFID (2012), 'TDP Business case'.

DFID (2013), 'TDP Annual Review 2013'.

McCormick, B., (2013), TDP M&E strategy

TDP, (2013), 'TDP Revised workplan'

TDP (2014), 'TDP Inservice Teacher Training Strategy'.

TDP (2014), 'Teacher Development Team selection criteria.'

TDP (2014), 'TDP TERMS OF REFERENCE: EDOREN SERVICE PROVISION for TDP IN-SERVICE BASELINE STUDIES and SUPPLY AND DEMAND (THEMATIC) STUDIES', draft.

Vogel, I., (2012), 'Review of the use of theory of change in international development', Review paper for the Department of International Development. DFID.

White, H., (2009), 'Theory-based impact evaluation: principles and practice', 3ie Working Paper 3. International Initiative on Impact Evaluation.

Annex B Intervention factsheets

B.1 Output 1: In-service training component

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

Intervention	In-service training	
1. Current alignment with overall programme design		Source
Output Logframe	Output 1: Improved in-service training of primary and junior secondary school teachers	TDP Logframe, May 2014
Output indicators:	<ul style="list-style-type: none"> 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 'Teacher Facilitators' trained to support teachers in-service training: this role provides the support package to teachers 	TDP Logframe, May 2014
Budget Allocated	DFID: £24.9m (60%) [Govt: £16.4m; hence total: £41.3m]	TDP VFM Strategy Final, Feb 2014
Impact Weighting	80%	TDP Logframe, May 2014
2. Intervention logic		
Objectives	<ul style="list-style-type: none"> • 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 educators responsible for the coordination, development and delivery of in-service programmes • Assist states to provide regular and on-going training and support to a total of 62,000 teachers by 2019 	TDP INSET Strategy, Jan 2014
Intended Outcomes	<ul style="list-style-type: none"> • More effective teachers in target schools • More effective teacher educators in colleges 	TDP Logframe, May 2014
Outcome indicators	<ul style="list-style-type: none"> • Percentage change (year on year) in the proportion of time teacher involves students in positive interaction 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	<ul style="list-style-type: none"> • Improved student learning in target schools 	TDP Logframe, May 2014
Impact indicators	<ul style="list-style-type: none"> • Percentage change (year on year) in student learning outcomes: +3 percentage points in each of English, Math, Science at the end of Phase 1 (measured through baseline in year 3) and another +3 percentage points at the endline of phase 2 (Year 5). Baseline to be established in October 2014 (through a school survey) 	TDP Logframe, May 2014
Beneficiaries	<ul style="list-style-type: none"> • Ultimate beneficiaries: Students in P1-6 and JSS 1-3 • Primary beneficiaries: Teachers in P1-6 and JSS 1-3 of English (31,000), Math (31,000), Science & Technology (31,000) 	TDP INSET Strategy, Jan 2014

	<ul style="list-style-type: none"> Secondary Beneficiaries: State-based teacher educators (teacher development teams) and school supervisors (teacher facilitators) 	
3. 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: For INSET programme implementation, the proposal is to implement a two year pilot with 6000 teachers in 1500 across all LGEAs in the 3 focus states.	
Geographical Scope	Whole programme: Northern Nigeria, in 6 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 6000 teachers from 1500 schools across all 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 continue till end of the programme.	
4. 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, Jan 2014
'Trainer in the pocket':	Access to audio-visual resources, anytime and anywhere, ensured through the use of mobile technology;	TDP INSET Strategy, Jan 2014
Continuous support	Ensured continuous support to teachers over a few years through multiple layers and mechanisms (workshops, cluster meetings, classroom observation, peer support, self-study materials etc.) instead of one-off trainings. The support mechanism for continuous professional development institutionalised within the school (teacher development team) and state (teacher facilitators).	TDP INSET Strategy, Jan 2014
5. Intervention details <i>sequence of these activities is not always clear from the docs we have</i>		
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 Number of 6 teachers per school in English, Maths and Science & 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 HT activities	Engage the HTs 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 Teacher Facilitators (at LGEA and cluster level). In consultation with SUBEB and CoEs, recruit a pool of teacher educators as the Teacher Development Team based in the state. Recruit Teacher Facilitators from the current School Supervisor cadre (SSOs/QAOs) at LGEA level based on a rigorous selection process	
Material development	<ul style="list-style-type: none"> 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 	

	<p>and Head teacher training materials from both ESSPIN and GEP 3.</p> <ul style="list-style-type: none">TDP will then also develop audio visual materials for use by all the beneficiary groups of the Teacher Development Team, the Teacher Facilitators 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 & Technology, then JS 1-3 English, Maths and Science & 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 & technology	Programme aims to undertake a technology assessment study to decide on final options for the mobile technology to support teachers. State govts. have communicated interest in supporting procurement of technology tools to aid teachers 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.	
6. Intervention stakeholders		
Implementing partners	DFID (through Mott MacDonald) State in three focal states: SUBEB Federal: NCCE	TDP Annual Review Final March 2014; TDP INSET Strategy, Jan 2014
Evidence partners	EDOREN	
Funding partners	DFID, State Governments	
7. Intervention-specific M&E plans		
Internal	M&E strategy doc; what else?	
External	External team to implement baseline survey for in-service and subsequent evaluations	
8. 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 EiA's success be transplanted to the Nigerian context and to what extent?		Annual Review 2014
9. Directly influencing interventions		
ESSPIN	In Jigawa, ESSPIN is also running teacher/HT training	Revised GEP3 log frame, 2013
GEP3	In Katsina and Zamfara, GEP3 is supporting girls' education	
USAID's P 1/2 reading skills in Hausa project	Sokoto, Bauchi (will they intervene in any TDP states eventually?)	Annual Review 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, Jan 2014
Political context	<ul style="list-style-type: none">College of Education staff was on strike from Jan to end-Feb 2014. Was this a one-off or will this repeat itself?	TDP Annual Review Final

	<ul style="list-style-type: none"> 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 momentum? 	March 2014; TDP INSET Strategy, Jan 2014
10. Reference documents		
Intervention documents	TDP Logframe, May 2014 TDP INSET Strategy, Jan 2014	
Evidence documents (studies, evaluations)	TDP Annual Review Final March 2014 TDP Evaluation of First and On-going Activities Jan 2014	

B.2 Output 2: Pre-service training component

Intervention	Pre-service training	
11. Current alignment with overall programme design		Source
Output Log frame	Output 2: More effective teacher educators of primary and junior secondary school student-teachers	TDP Logframe, May 2014
Output indicators:	<ul style="list-style-type: none"> Number of CoE lecturer attending skills updating sessions, expressed in person-course-hours Percentage change in amount of support received by student-teachers from teacher-educators during teaching practice measured in whole time equivalent days Number of lecturer/student-teacher contact hours 	TDP Logframe, May 2014
Budget Allocated	DFID: £11.1m (100%)	TDP VFM Strategy Final, Feb 2014
Impact Weighting	10%	TDP Logframe, May 2014
12. Intervention logic		
Objectives	<ul style="list-style-type: none"> To support the implementation of the new NCE curriculum To facilitate the restructuring of CoEs consistent with the new NCE curriculum To support the enhancement of quality assurance processes and procedures To support NCCE QA processes and procedures To enhance active learning pedagogies and update subject knowledge of CoE staff To enhance quality of teaching practice 	Pre-service strategy (confidential)
Intended Outcomes	<ul style="list-style-type: none"> More effective teachers in target schools More effective teacher educators in colleges 	TDP Logframe, May 2014
Outcome indicators	<ul style="list-style-type: none"> Number of colleges scoring ‘good’ or ‘satisfactory’ in each of the five key areas of the QA assessment’ Percentage increase (year on year) in the level of student teacher satisfaction with the support received during teaching practice 	TDP Logframe, May 2014
Intended Impact	<ul style="list-style-type: none"> Improved student learning in target schools 	TDP Logframe, May 2014
Impact indicators	Percentage change (year on year) in student learning outcomes: +3% in each of English, Math, Science at the end of Phase 1 (measured through endline in year 3) and another +3% at the endline of phase 2 (Year 5).	TDP Logframe, May 2014

		(see logframe for milestone by phases)
Beneficiaries	<ul style="list-style-type: none"> - Ultimate beneficiaries: Students in P1-6 and JSS 1-3 - Primary beneficiaries: CoE lecturers and state-based teacher-educators - Secondary Beneficiaries: Teachers in P1-6 and JSS 1-3 of English, Math, Science & Technology 	TDP Logframe, May 2014; Revised BC
13. Scope of intervention		
Target Population	<ul style="list-style-type: none"> - No. of academic staff to be trained in Activity-based Learner-centred pedagogy: phase 1 is 363, phase 2 is 453 - No. of students-teachers to be trained to use "The Trainer-in-the Pocket" model: phase 1 is 1284, phase 2 is 2716 	Pre-service strategy (confidential)
Geographical Scope	CoEs in in the TDP support States of Jigawa, Katsina, and Zamfara in the first phase of TDP (2014-16) and then expand to Kaduna, Kano and Niger States during phase two (2016-19). The ultimate intention is to encourage NCCE to replicate the models, modalities, materials and training implemented in the six states to other colleges of education in Nigeria	Pre-service strategy (confidential)
Implementation Timeline	<ul style="list-style-type: none"> - No. of academic staff to be trained in Activity-based Learner-centred pedagogy: phase 1 is 363, phase 2 is 453 - No. of students-teachers to be trained to use "The Trainer-in-the Pocket" model: phase 1 is 1284, phase 2 is 2716 	Pre-service strategy (confidential)
14. Intervention components		
To support the implementation of the new NCE curriculum	<ul style="list-style-type: none"> - Dissemination of the curriculum in 5 COEs - Re-structuring of COEs from subject-based structure to one based on areas of specialization (ECCE, Primary, JSS, ANFE & Special Educ.) 	Pre-service strategy (confidential)
Building the capacity of teacher educators in pedagogy	<ul style="list-style-type: none"> - 816 teacher educators to be trained in activity-based pedagogy 	Pre-service strategy (confidential)
More opportunities for Teaching Practice	<ul style="list-style-type: none"> - "Associate schools" created. - Improved supervision and guidance to student teachers - 4000 student teachers trained in the use of "Trainer-in-the Pocket" model 	Pre-service strategy (confidential)
To support the enhancement of quality assurance processes and procedures	<ul style="list-style-type: none"> - Review and improvement of QA Processes and Procedures at COEs and NCCE - Dissemination of QA Tool Kit - Self-assessment by COEs - Accreditation of COEs by NCCE 	Pre-service strategy (confidential)
15. Intervention details (inputs/activities)		
Organise workshops for academic staff on the new Curriculum, college re-structuring, QA toolkit	TDP will support NCCE to organise state-based awareness raising events initially in the three TDP support states (Jigawa, Katsina and Zamfara) for key government and COEs officials and this will later be extended to the remaining three states (Kaduna, Kano and Niger).	Pre-service strategy (confidential)
Adapt/develop materials and develop capacity on pedagogy		Pre-service strategy (confidential)
Adapt/develop student materials supporting the trainer in the pocket		Pre-service strategy (confidential)

Create Associate schools		Pre-service strategy (confidential)
Create COE-SUBEB linkages		Pre-service strategy (confidential)
Train TP supervisors and mentors		Pre-service strategy (confidential)
Assist CoEs to develop and implement plans for restructuring consistent with the new curriculum		Pre-service strategy (confidential)
Build QA Capacity and support self-assessment by COEs		Pre-service strategy (confidential)
To support the implementation of the new NCE curriculum		Pre-service strategy (confidential)
To facilitate the restructuring of CoEs consistent with the new NCE curriculum		Pre-service strategy (confidential)
To support the enhancement of quality assurance processes and procedures		Pre-service strategy (confidential)
16. Intervention stakeholders		
Implementing partners	DFID (through Mott MacDonald) State in three focal states: SUBEB/CoE	TDP Annual Review Final March 2014
Evidence partners	EDOREN	
Funding partners	DFID	
17. Intervention-specific M&E plans		
Internal	M&E strategy doc; what else?	
External	EDOREN to propose a suitable evaluation approach and method in the evaluation framework.	
18. Identified assumptions and risks		
It is being assumed that NCCE will implement the use of the QA instrument in CoEs which TDP aims to review and improve	Comments:	TDP Logframe, May 2014
TDP pre-service package leads to better educators, and thus better NCE certified teachers and this results in improved learning outcomes	Comments:	TDP Logframe, May 2014
There is political will in the college of	Comments:	TDP Logframe, May 2014

education to restructure.		
Issues around teacher workforce management start to be addressed by SUBEBs and constraints begin to be tackled effectively	Comments:	TDP Logframe, May 2014
Corruption in the education system begins to be addressed especially in relation to teacher management issues	Comments:	TDP Logframe, May 2014
Security situation in the North does not affect lectures and student attendance in college and schools	Comments:	TDP Logframe, May 2014
Security situation in the North does not affect teacher and student attendance in school	Comments:	TDP Logframe, May 2014
19. Directly influencing interventions: assumptions and risks		
Synergies with in-service	“Trainer in pocket”: The model and its associated technology and materials will be adapted and used as part of the capacity building programme of the pre-service component. Similarly, as part of its attempt to improve the effectiveness of Teaching practice, the pre-service component will support colleges and SUBEBs to establish “Associate schools” in all the three states. It is planned that such schools will mainly if not exclusively be those the in-service component is already working with. This will facilitate the sharing of materials but more importantly gives the two components the opportunity to draw from each other’s strengths and successes, learn from their respective challenges and try out new ideas and practices in more or less the same setting and thereby reinforce their respective activities.	Pre-service strategy (confidential)
NCCE’s own reforms	National Commission for Colleges of Education (NCCE) has designed its own strategy for reform.	Pre-service strategy (confidential)
Political context	<ul style="list-style-type: none"> - College of Education staff were on strike from Jan to end-Feb 2014 - Is there enough political will to reform CoEs? - Will there be “slippage” in meeting programme milestones as a result of the general elections momentum? 	TDP Annual Review Final March 2014
20. Reference documents		
Intervention documents	TDP Logframe, May 2014 Pre-service inception report (confidential)	
Evidence documents (studies, evaluations)	TDP Annual Review Final March 2014 TDP Evaluation of First and On-going Activities Jan 2014	

B.3 Output 3: Results & Evidence (R&E) component

Intervention	Research & Evidence	Source
21. Current alignment with overall programme design		
Output Logframe	Output 3: Evidence-based research informing decisions on teachers' effectiveness and efficiency	TDP Logframe, May 2014
Output indicators:	<ul style="list-style-type: none"> d) Number of studies completed in the year e) Number of studies published in the year f) Use of evidence from studies by policy makers 	TDP Logframe, May 2014 (see logframe for milestone by phases)
Budget Allocated	DFID: £1.1mn	TDP VFM Strategy Final, Feb 2014
Impact Weighting	10%	TDP Logframe, May 2014
22. Intervention logic		
Objectives	<ul style="list-style-type: none"> To demonstrate program results To provide evidence to support the theory of change To monitor program activities 	TDP Inception Report Feb 2014
Intended Outcomes	<ul style="list-style-type: none"> 	
Outcome indicators	<ul style="list-style-type: none"> 	
Intended Impact	<ul style="list-style-type: none"> Improved student learning in target schools 	TDP Logframe, May 2014
Impact indicators	<ul style="list-style-type: none"> Percentage change (year on year) in student learning outcomes: +3% in each of English, Math, Science at the end of Phase 1 (measured through endline in year 3) and another +3% at the endline of phase 2 (Year 5). 	TDP Logframe, May 2014 (see logframe for milestone by phases)
Beneficiaries	<ul style="list-style-type: none"> Policymakers in the education sector in Nigeria – e.g FME, SMOE, TDP for improved implementation 	
23. Scope of intervention		
Target Population	<ul style="list-style-type: none"> the teachers and head teachers in classrooms and schools; the TFs working with teachers and head teachers; the TDTs in each state working with the TFs, and TDP (using national and international experts) working with TDTs (see Figure 1, below). 	TDP Results and Evidence Strategy May 2014
Geographical Scope		
Implementation Timeline	Year 1 Implementation Quarterly Deliverables <ul style="list-style-type: none"> Q1: Instruments developed for in-service and pre-service (MLA Science & Technology, TDNA S&T, Classroom observation, teacher motivation survey, and pre-service student teacher survey) 	TDP Inception Report Feb 2014

	<ul style="list-style-type: none"> Q2: Pilot Baselines completed Q3: Field work for in-service and pre-service baselines completed (MLA, TDNA, classroom observation, teacher motivation, pre-service student teacher, pre-service case study) Q4: PMIS developed and operational 	
24. Intervention components		
The evaluation of the Programme in terms of its impact, outcomes and outputs.	This consists of studies that focus on logframe reporting and a combination of qualitative and quantitative methods for outcome evaluation of both in-service and pre-service treatments	TDP Results and Evidence Strategy May 2014
Studies that evaluate the results of the Programme but which are intended to add to the educational evidence base in Nigeria.	Dealing with evidence gaps identified in the Business Case, and reinforced by the EDOREN literature review (EDOREN, 2013) either as a consequence of evaluation of the intervention components or the focus of specific studies directed at evidence gaps	TDP Results and Evidence Strategy May 2014
The monitoring of Programme activities to ensure the quality of their implementation.	<p>There are three basic kinds of monitoring:</p> <ol style="list-style-type: none"> 1. Of TDP activities; i.e. workshops to train teacher educators who will train student teachers using simple questionnaires and evaluation forms 2. Of Classroom activity by teachers taking part in TDP using classroom observation visits; building on existing school monitoring and QA systems 3. Of Pre-service colleges: TDP staff in the state teams carry out monitoring visits to colleges and to use simple checklists to register what is happening. 	TDP Results and Evidence Strategy May 2014
25. Intervention details		
In-service monitoring	<p>Quantitative includes the following design options:</p> <ul style="list-style-type: none"> a comparison of control and treatment groups through a quasi-experimental approach using randomised sampling (is the most practical and effective approach) pre-/post-test (i.e. no control group); (not ideal, will monthly be used if other designs cannot be implemented) Treatment 1 (ESSPIN/GEP) v treatment 2 (TDP). (This measures the added value of TDP where ESSPIN and GEP are working in the same schools. <p>Qualitative approach includes School and school cluster case studies</p>	TDP Results and Evidence Strategy May 2014
Pre-service monitoring	<ul style="list-style-type: none"> Case studies, given the small sample size of CoEs involved in intervention and the need to capture context around restructuring of CoE Quantitative: Teacher- student survey to collect data on contact hours, lecturer support to student teachers; student teacher satisfaction 	TDP Results and Evidence Strategy May 2014
Results and Evidence	<ul style="list-style-type: none"> Evidence as a consequence of the evaluation of Components 1 & 2- a combination of logframe reporting and qualitative explanatory studies Thematic studies to fill evidence gaps (e.g study of motivation of general teaching force) to be carried out after the first year of implementation. 	TDP Results and Evidence Strategy May 2014

	<ul style="list-style-type: none">Technology Environment Assessment: A survey to capture the nature and use of existing technologies by teachers and schools to enable planning of in-service audio-visual material for teachers	
Processes	<ul style="list-style-type: none">Development Advisory Team consisting of EDOREN, GEP and ESSPIN; representatives from Federal (FME, UBEC AND NCCE) and state levels.Instrument development in 1-3 states, piloting in different states to enable trial of the process	TDP Results and Evidence Strategy May 2014
Sampling	For design of evaluation of Pre and In-service training respectively, baseline of quasi-experimental studies will require: <ul style="list-style-type: none">complete details of schools, of all the schools, head teachers, teachers and students by state and LGA and primary and junior secondary.To be clear what schools ESSPIN and GEP are working at implementation stage of TDP.	TDP Results and Evidence Strategy May 2014
Instruments	All the instruments are based on the logframe indicators. The following instruments are available currently: <ul style="list-style-type: none">English & mathematics MLA at P2/P4 (ESSPIN 2010 baseline; GEP equivalents 2013/4).TDNA English & mathematics for Primary teachers for P1-6 and JSS 1-3 (ESSPIN 2010 baseline & GEP 3 2013/4).Classroom observation (ESSPIN 2010 baseline).¹¹ Those instruments that may be needed: <ul style="list-style-type: none">'MLA' Science and technology at P2/P4 & JSS 3¹²Teacher motivation surveyTDNA Science and technologyStudent teacher satisfaction and experience of teaching practice support.	TDP Results and Evidence Strategy May 2014
26. Intervention stakeholders		
Implementing partners	SMoE, SUBEB, LGEA	TDP Business Case July 2012
Evidence partners	EDOREN, UNICEF, GEP3, ESSPIN	
Funding partners	DFID Nigeria	
27. Intervention-specific M&E plans		
Internal		
External		
28. Identified assumptions and risks		
Politico- security situation	The lead-up to the 2015 elections in Nigeria implies an increase in election-related activities which may disturb planned R&E activities (piloting instruments and conduct of baselines)	TDP Inception Report
Strikes	State Colleges of Education are prone to strikes	TDP Inception Report
Crime	High crime rates especially in Zamfara	TDP Inception Report
29. Directly influencing interventions		
ESSPIN	In Jigawa, ESSPIN is also running teacher/HT training	
GEP3		
30. Reference documents		

¹¹ These may be GEP observation instruments but this is not clear at the moment.

¹² It is not clear if any student outcome measures are required at P6 level.

Intervention documents	TDP Logframe, May 2014 TDP Research & Evaluation Strategy	
Evidence documents (studies, evaluations)		

Annex C The TDP logframe (version October 2014) and research and evidence strategy

This is the logframe as provided by the TDP in October 2014.

TDP Logical framework¹³

PROJECT NAME	TEACHER DEVELOPMENT PROGRAMME – Nigeria					
IMPACT Improved student learning in target schools	Impact Indicator Percentage change in knowledge ¹⁴ and comprehension of three core subjects of primary school students in level 2 and 5 ¹⁵		Phase 1 States			
			Baseline Yr 1 (2014)		Endline Yr 4 (2017)	
		Planned	English Maths Science		English Maths Science	+3% (TBD) ¹⁶ +3% +3%
		Achieved	English Maths Science		English Maths Science	
		Source: Results from TDP intervention schools using ESSPIN/TDP MLAs				

¹³ Each indicator will be reported separately by State; where figures are given it will be noted which state data are used

¹⁴ From the P4 S&T test specification, the various questions types are identified: state, describe, compare, explain and predict. In Bloom's taxonomy, state and describe are associated with knowledge while compare, explain and predict are associable with comprehension. The S&T P2 test focuses largely on knowledge as while the ESSPIN P2 English and Math tests focuses entirely on knowledge.

¹⁵ To be disaggregated by (a) Primary 2, Primary 5 and (b) by gender and (c) by region

¹⁶ To be determined after completion of the baseline

OUTCOME 1	Outcome 1 Indicators		Phase 1 States							
			Baseline Yr 1 (2014)			Midline Yr 2 (2015)		Endline Yr 4 (2017)		
More effective teachers in target schools	Outcome Indicator 1.1 Percentage change in the proportion of time teacher involves students in positive interaction in a lesson ¹⁷	Planned				5%		+5%		Assumptions 1. More effective teaching results in improved learning outcomes 2. Improvements in the management of schools reinforces teachers’ and students’ performance 3. (In)Security situation and or elections do not affect teacher and student attendance in school
		Achieved								
		Source: Classroom observation (TDP 2014 classroom observation instrument)								
	Outcome Indicator 1.2 Change in teacher absenteeism as a percentage of the average days absent to the number or contracted working days measured year on year ¹⁸	Planned				-2%		-2%		
		Achieved								
		Source: TDP Teacher Interview								
	Outcome Indicator 1.3 Percentage change in teacher subject knowledge and competency ¹⁹	Planned	English		English	+5%	English	+5%		
			Maths		Maths	+5%	Maths	+5%		
			Science		Science	+5%	Science	+5%		
		Achieved	English		English		English			
Maths				Maths		Maths				
Science		Science		Science						
Source: TDNA scores from TDP studies										

¹⁷ To be disaggregated by (a) Primary 2, Primary 5 and (b) by gender and (c) by region

¹⁸ To be disaggregated by (a) Primary 2, Primary 5 and (b) by gender and (c) by region

¹⁹ To be disaggregated by (a) Primary 2, Primary 5 and (b) by gender and (c) by region

OUTCOME 2	Outcome 2 Indicators		Phase 1 States				Phase 2 States				Assumptions	
			Baseline Yr 1 (2014)		Endline Yr 4 (2016)		Baseline Yr 4 (2016)		Endline Yr 4 (2018)			
More effective teacher educators in colleges	Outcome Indicator 2.1 Number of colleges scoring ‘good’ or ‘satisfactory’ in each of the five key areas of the QA assessment ²⁰	Planned									1. NCCE implements the use of the QA instrument in Colleges 2. There is political will in the CoEs to restructure. 3. (In)Security situation and or elections do not affect lectures and student attendance in CoEs 4. Strike by CoEs or lecturers does not affect lectures or student attendance.	
		Achieved	Leadership									
			Curriculum									
			Infrastructure									
		Assessment										
		Student Sppt										
	Source: CoEs’ self-assessment using NCCE QA toolkit.											
	Outcome Indicator 2.2 Percentage change in amount of support received by student teachers from teacher educators during teaching practice measured in whole time equivalent days	Planned										
		Achieved										
		Source: Student survey conducted by TDP, Observations of teaching practice support										
	Outcome Indicator 2.3 The extent to which lecturers use activity-based, learner-centred approaches to implement the curriculum	Planned										
		Achieved										
		Source: Student survey conducted by TDP, Observations of teaching practice support										

²⁰ These key areas are: Leadership and management and organisation; Curriculum organization and implementation; infrastructure and learning resources; assessment and evaluation; students' support and progression.

OUTPUT 1	Output 1 Indicators		Phase 1 states		Phase 1 and 2 States			Assumptions
			Baseline	Milestone (Yr2)	Milestone (Yr3)	Milestone (Yr4)	Target (Yr5)	
Improved in-service training of primary and junior secondary school teachers	Output Indicator 1.1 Number of teachers trained in the year ²¹	Planned (PR)						1. State Governments’ contributions to TDP are made in full and on time 2. Issues around teacher workforce management start to be addressed by SUBEBs and constraints begin to be tackled effectively 3. Corruption in the education system begins to be addressed especially in relation to teacher management issues RISK RATING = MEDIUM
		English	0	3144	8960	14776	20592	
		Maths	0	3144	8960	14776	20592	
		Science ²²	0	1008 ²³	3144	8960	14776	
		Planned (JSS)	0	0	3816	3816	3816	
		English	0	0	3816	3816	3816	
		Maths	0	0	3816 ²⁴	3816	3816	
		Science						
		Achieved (PR)						
		English						
		Maths						
		Science						
		Achieved (JSS)						
		English						
		Maths						
	Science							
Source: Programme records								
Output Indicator 1.2 Number of head teachers in TDP intervention schools trained in leadership and management of staff (person	Planned (PR)	0	1572	4480	7388	10296		
	Planned (JSS)	0	0	1908	3816	5724		
	Achieved							

²¹ To be further disaggregated by (a) Primary 1-6 and JSS levels (b) gender and (c) region. Total number is 20592 + 20592 + 6(3816) = 64 080

²² These are not new teachers, i.e. the numbers are taken from Maths and English teachers. They are not to be counted as part of the overall total of teachers trained.

²³ Not to be counted as part of the overall total as these are taken from the Math and English teacher numbers

²⁴ Similar to the calculations for primary teachers, this is not to be counted as part of the overall total as teachers are taken from the Math and English figures

	days) ²⁵	(PR)						
		Achieved (JSS)						
		<i>Source: Programme records</i>						
IMPACT WEIGHTING (%) 80	Output Indicator 1.3 Number of 'Teacher Facilitators' trained to support teachers in-service training ²⁶	Planned	0	262	864	1466	2068	
		Achieved						
		<i>Source: Programme records</i>						

²⁵ To be disaggregated by (a) Primary and JSS (b) gender and (c) region

OUTPUT 2	Output 2 Indicators		Phase 1 states		Phase 1 and Phase 2 States			Assumptions [These numbers at milestones will have to be replaced to reflect not the number of staff, but the number of person-course-hours.
			Baseline	Milestone (Yr2)	Milestone (Yr3)	Milestone (Yr4)	Target (Yr5)	
More effective teacher educators of primary and junior secondary school student teachers	Output Indicator 2.1 Number of CoE lecturer attending skills updating sessions, expressed in person-course-hours	Planned	0	363	816	816	816	RISK RATING = MEDIUM
		Achieved						
		Source: PMIS						
IMPACT WEIGHTING (%) 10	Output Indicator 2.2 Number of lecturer-student teacher contact hours	Planned	TBD	10%	10%	10%	10%	
		Achieved						
		Sources: Programme records, Analysis of teaching schedules, Observations (sampled), Survey of student teachers						

OUTPUT 3	Output 3 Indicators		Phase 1 states		Phase 1 and Phase 2 States			Assumptions	
			Baseline	Milestone (Yr2)	Milestone (Yr3)	Milestone (Yr4)	Target (Yr5)	1. There is political will to implement recommendations from studies	
Evidence-based research informing decisions on teachers' effectiveness and efficiency	Output Indicator 3.1 Number of studies completed in the year	Planned	0	2	3	4	5	RISK RATING = LOW	
		Achieved							
	Source: Programme records								
	Output Indicator 3.2 Number of studies published in the year	Planned	0	1	2	3	5		
		Achieved							
IMPACT WEIGHTING (%) 10	Output Indicator 3.3 Use of evidence from studies by policy makers	Source: Programme records							
		Planned	0	0	1	2	3		
		Achieved							
Source: Programme commissioned study, Document analysis, Interviews with policy makers and associated staff									

C.1 TDP Research and evidence strategy

251. The TDP R&E strategy (McCormick 2013) has two elements:

- Monitoring of activities at the operational level of the programme through continuous data collection and analysis with the view to improving quality.
- Evaluating the impact and outcomes of the programme, and providing an evidence base for strategic decision making on teacher efficiency and effectiveness.

C.1.1 Monitoring

252. The proposed TDP monitoring takes two forms. The first is monitoring of TDP activities such as teacher training workshops, through simple questionnaires given to participants, to check that these activities happen, and that (e.g.) teachers participate, and provide some structured and unstructured feedback.

253. The second type of monitoring focuses on the classroom activity of teachers who participate in the TDP. This will be done through classroom observations that build on existing school monitoring and QA systems. Variations of the current ESSPIN model which uses head teachers and School Services Officer as prime monitors of teacher classroom practise will also be explored. For pre-service training, monitoring will take the form of visits by TDP state staff to the Colleges of Education.

254. The TDP evaluation framework set out in the present framework proposes to use these monitoring data to support rigorous factual analysis across the TDP's causal chain (see section 4.2).

C.1.2 Evaluation

255. The proposed approach to evaluation in the TDP R&E strategy is more detailed. The evaluation approach proposed in the present document should be congruent with the R&E strategy approach, and some elements of the evaluation will extend or adjust the R&E approach, and others are identical to what is proposed.

256. As proposed in the R&E strategy, evaluation has two aspects. The first focuses on the logframe indicators in order to measure progress and final target impact and outcomes of the programme. The second element provides evidence on the efficiency and effectiveness of teachers, to fill gaps in literature and provide evidence for decision makers in teacher education. As per the R&E strategy, the proposed design of the evaluation will balance between the two main elements of the evaluation: 'reporting on the log frame' and 'giving explanations'; and also incorporate different approaches for the in-service and pre-service components. The approach to evaluation proposed in the current document shifts the balance towards 'giving explanations', or in our terminology, 'learning', while also expecting to report on log frame indicators. This is possible because DFID have additional evaluation resources available for the evaluation of TDP.

257. The R&E approach to evaluating the in-service component recommends a quantitative approach that is 'quasi-experimental...with suitably randomised sampling', possibly complemented by case studies of schools to explore mechanisms of change (McCormick 2013: 11-12). This quasi-experimental and case study approach is proposed in the present document.

258. The R&E approach to evaluating the pre-service component recommends case studies of a small number of Colleges of Education, largely using qualitative data but complemented with a student-teacher survey to gather information on contact hours, lecturer support and student teacher satisfaction. These

case studies will be undertaken by the TDP, but the pre-service component is not included in the evaluation framework.

259. In addition to these generalised approaches, the R&E strategy sets out specific ‘evidence-focused studies’ that attempt to provide additional learning on the mechanisms of change for the TDP and that fill gaps in the evidence base. These should be congruent with the proposed EDOREN studies on teaching (currently including one study on teacher supply and demand in Katsina that is in draft, and a planned study on teacher management and deployment). In the December 2013 version of the R&E strategy, these studies are presented tentatively and will be finalised in the first year of implementation, but could include studies on teacher motivation, or a tracer study on student teachers upon graduation. The current evaluation framework does not set out these types of studies in detail, but would seek to incorporate their findings given their likely relevance, and the evaluation and wider EDOREN team will seek to continue a dialogue with the TDP R&E team on their subject and scope.

C.1.3 Instruments and logframe

260. The R&E strategy discusses relevant instruments and their relation to the logframe (see Annex C). The current evaluation framework proposes some changes to what is currently contained in these two documents (the reasons for this are outlined in more detail in section 0). In this section, we note what is currently contained in these documents and summarise the changes proposed. One observation from the R&E strategy that will be picked up later is that the language of testing for students and teachers is likely to be extremely important – because tests in English will almost certainly lead to measuring lower outcomes.

C.1.3.1 Impact measures

261. The May 2014 logframe (D4) impact indicator is percentage change (year on year) in student learning outcomes, disaggregated by a) primary 1-3, primary 1-6 and JSS, b) by gender, and c) by region. The logframe anticipates a 3% increase by year 3 on a baseline of 11-12% of students at adequate learning levels (taken from the ESSPIN 2010 survey at p4), and measurement at year 1 and year 3 for the TDP phase 1 states, and year 3 and year 5 for the TDP phase 2 states. The R&E strategy (which refers to logframe D3 so is not entirely up to date with the logframe) proposes to measure this using existing instruments used by ESSPIN:

- An MLA to measure English and mathematics at p2 and p4
- A TDNA for English and mathematics for primary teachers teaching grades 1-6 (for outcome indicators on teacher quality and effectiveness)
- Classroom observations (for outcome indicators on teacher quality and effectiveness).

262. The TDP R&E strategy notes that this implies that further instruments are required:

- An MLA for science and technology at p2 and p4
- A TDNA for science and technology (for outcome indicators on teacher quality and effectiveness)
- Teacher motivation survey (for a thematic study)
- Student-teacher satisfaction and experience of teaching practice support (to assess pre-service outcome on more effective teacher educators in colleges)

263. The R&E strategy proposes two possible alternatives to the development of these tests to measure the impact indicator on learning: using p6 public examinations, or not testing at every grade (McCormick 2013: 16). The evaluation framework here proposes a slightly different solution, which is designed to maximise the cost effectiveness of measuring impact on student learning. Described in more detail in section 4.1.3, this involves tracking a ‘panel’ or ‘cohort’ of students as they progress from p2 to p5 over three years (2014 to 2017), moving through classes where teachers are exposed to the TDP’s in-service component (p1-3 in 2014/15 and p4-6 in 2015/16/17) as a treatment, and with a control of similar students

in schools where the TDP is not operating. This means that the logframe will report percentage change in learning outcomes, but not specifically at p2 or p4 but for a specific group of children. This approach is taken because it requires a much smaller sample size to detect an impact of 3% at the State-level. Given resource constraints, it is not currently proposed to measure learning outcomes in the TDP phase 2 states. The evaluation will also seek to measure learning in terms of changes of examination scores at p6, though this remains subject to some concerns about reliability and the ability of these scores to differentiate small changes in levels of learning, and they are unlikely to be appropriate to compare results over time given changes to exam process.

264. This means that the proposed evaluation framework has the following differences from the logframe, at impact level:

- Learning outcomes will be measured in the TDP phase 1 states only, not both as in the D4 logframe.
- The baseline will be in year 1 (2014) and the endline in year 4 (2017), not in year 3 as in the D4 logframe.
- Change in learning outcomes will not be disaggregated by p1-3, p1-6 and JSS, because they will refer to a single cohort of students (and these students will not be at JSS at all).
- Change in learning outcomes may be disaggregated by gender (and then not within each state), but this depends on the magnitude of the change (only a larger than expected change will be detected) and the difference between effects of boys and girls (only a large difference will be detected).
- Changes in learning outcomes are unlikely to be disaggregated by region if this means regions within different states, for the same reasons. They should be able to be disaggregated by state, unless changes are smaller than expected.
- These last four points on timing and disaggregation apply to all indicators collected by the proposed quantitative surveys on the impact and outcome of in-service training.
- Changes in learning outcomes will be attributable to the in-service output, not the pre-service.

C.1.3.2 Outcome measures

265. The first logframe outcome (around in-service training) is more effective teachers in target schools, measured by three indicators:

- i. proportion of time teacher involves students in positive interaction in a lesson, expected to increase by 5% after two years to be measured by classroom observations using an instrument similar to the ESSPIN 2010 baseline, but which, according to the R&E strategy, will need to be refined. The proposed evaluation framework will measure this, with the differences noted under impact above.
- vii. teacher absenteeism as a percentage of the average days absent to the number or contracted working days measured year on year, to reduce by 2% every two years, and measured by SUBEB administrative records and monitoring/school data directly collected. As the R&E strategy notes, it is unlikely that the SUBEB administrative records will everywhere be of sufficient reliability to collect this indicator, and so the quantitative surveys proposed in this evaluation framework will measure this. The differences noted under impact above apply.
- ii. change (year on year) in teacher subject and pedagogic knowledge competency, to improve by 5% every two years as measured by TDNA studies for English, Mathematics and Science&Technology. The R&E strategy proposes using the ESSPIN 2010 instrument and adding an instrument for S&T, considering the use of video and adding qualitative studies. The quantitative surveys proposed in this evaluation framework would use this approach (without video), though with the caveats listed under impact above.

266. The second logframe outcome (around pre-service training) is more effective teacher educators in colleges, measured by two indicators:

- i. Number of colleges scoring 'good' or 'satisfactory' in each of the five key areas of the QA assessment, currently without a target, to be measured by the NCCE QA toolkit and accreditation

process. The R&E strategy notes that this will be cross-checked by researchers who carry out college case studies.

- ii. Percentage increase (year on year) in the level of student teacher satisfaction with the support received during teaching practice, to increase by 5% in two years, as measured by student surveys conducted by TDP and observations of teaching practice support. The R&E strategy notes that the observations will provide explanations for increases or not that are achieved.

267. As noted elsewhere, the evaluation will not address the pre-service component directly.

C.1.3.3 Output measures

268. There are three TDP outputs, corresponding to in-service, pre-service and results and evidence; here we focus on the first. Output 1 is improved in-service training of primary and junior secondary school teachers, measured by:

- i. Number of teachers trained in the year, measured by programme records.
- ii. Number of head teachers trained in management and development of staff (person days) measured by programme records.
- iii. Number of 'Teacher Facilitators' trained to support teachers' in-service training measured by the ESSPIN teaching and learning survey. The R&E strategy expects this indicator to be measured by programme records, and we do not think a survey would be a more appropriate data source for this.

Annex D Principles, standards and criteria guiding TDP's evaluation

269. According to the OECD (1991) principles for evaluation of development assistance, the two main purposes of an evaluation are:

- I. to improve future aid policy, programmes and projects through feedback of lessons learned;
- II. to provide a basis for accountability, including the provision of information to the public.

270. In order to fulfil these objectives the TDP evaluation must generate high quality and dependable information. This will be achieved by adhering to certain principles, standards and criteria for evaluations have over the years been defined by the international community.

D.1 DAC evaluation principles

D.1.1 Impartiality and Independence

271. Impartiality and independence are crucial throughout each stage of the evaluation process. The evaluation should be impartial and independent from the process concerned with policy making, delivery and management of the development assistance being evaluated. Impartiality increases the credibility of the evaluation and avoids bias in findings. Independence provides legitimacy to the evaluation and reduces potential conflicts of interest where programme implementers evaluate their own activities.

272. The institutional set up of the evaluation must be such to ensure that the evaluation is being carried out by those who are independent of the design and implementation of the processes being evaluated. Governance arrangements for management and quality assurance are key to maintaining independence and impartiality (DFID, 2013).

273. In the case of the TDP, there are some issues of independence that need to be noted, because EDOREN is designing this evaluation framework and conducting the evaluation, providing (through a different EDOREN workstream separated by Chinese walls) quality review of data collected as part of the surveys, is also responsible for TDP annual reviews, and has in the past provided comments on the TDP design and some advice to results and learning. 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 problems for the impartiality and independence of the TDP's evaluation.

274. Our approach to mitigating these problems is to ensure that 1) all evaluation outputs are peer reviewed, 2) data and analytical approaches for both quantitative and qualitative data are transparent (i.e. publicly available for scrutiny), and 3) to contract independent annual review teams as per EDOREN's inception report.

D.1.2 Credibility

275. The credibility of an evaluation is dependent on both the expertise and independence of the evaluators, and the degree of transparency of the evaluation process. Transparency throughout the evaluation process is crucial to establish credibility and legitimacy of the evaluation process. This includes keeping the evaluation process open by publishing data, processes and findings and to make the results publically available. It should be possible to revisit and reanalyse primary data that has led to evaluation findings (DFID, 2013). Evaluation reports must distinguish between findings and recommendations. Relevant information to support findings should be included in a way that does not compromise sources. (OECD, 1991).

276. As noted above, 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.

D.1.3 Usefulness

277. To be useful for decision-making, evaluation findings must be perceived as relevant and useful and be presented in a clear and concise manner. They must reflect the needs and interests of the relevant development partners and be easily accessible. Evidence from evaluations must also be timely so the evidence they generate are available at the time. Adhering to the DAC principle of utility DFID measures utility by –

- i. Timelines ensuring that the evaluation reports are produced at relevant points in the policy and programme cycle
- ii. Relevance: ensuring by asking the questions to which users want to know the answers
- iii. Quality – ensuring that the evaluation is credible, reliable and can be used for decision making

278. In this case, the main purpose of the evaluation is to provide learning for other contexts, so the provision of the main evaluation findings following the completion of TDP's contract is not so problematic. Our approach to ensuring the relevance of the evaluation questions is to consult widely around this document. High quality expertise will be used in conducting the evaluation to ensure credibility and reliability.

D.1.4 Participation of donors and recipients

279. Donors and participants should be involved in the evaluation as far as possible. Participation increases the quality of the evaluation and gives all parties concerned an opportunity for learning by doing, increasing long term sustainability after the donor has left.

280. TDP's evaluation aims to fulfil the principles of usefulness and participation above through strategic and sustained stakeholder engagement and information dissemination at various stages of the evaluation process. This is also part of TDP's Results and Evidence approach, and the evaluation's proposed approach to stakeholder engagement, for which see section 234.

281. This draft of the evaluation framework is intended for consumption by TDP and DFID Nigeria to validate the overall approach and the detail around the interventions. However, a final version should be shared with secondary users as well.

D.2 DFID Evaluation standards

D.2.1 Quality Standards

282. EDOREN will adhere to OECD's DACs quality standards for evaluation when designing and executing TDP's evaluation. Quality standards are related to every aspect of the evaluation: (1) Overarching (2) Planning and design and (3) Implementation and management.

D.2.2 Ethics

283. Evaluation must abide by relevant professional and ethical guidelines and codes of conduct for individual evaluators. Evaluators must be mindful of gender roles, ethnicity, ability, age, sexual orientation, language and other differences when designing and carrying out the evaluation (DAC Standards, 1994). DFID has drawn up its own principles for ensuring high standards of ethics in research and evaluations which will be adhered to as part of the TDP evaluation (DFID, 2011). These include –

- i. Researchers and evaluators are responsible for identifying the need for and securing necessary ethics approval for the study they are undertaking
- ii. Research and evaluation must be relevant and of the high quality with clear value
- iii. Researchers and evaluators should avoid harm to participants in studies
- iv. Participation in research and evaluation must be voluntary and free from any external pressure
- v. Ensure confidentiality of information, privacy and anonymity of participants
- vi. Respect cultural sensitivities
- vii. Adhere to UK human rights law
- viii. Publication and communication of all research and evaluation results
- ix. Research and evaluation must be independent of those implementing the programme
- x. Particular emphasis on participation from women and other social excluded groups

284. Conducting qualitative and quantitative field work 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). The evaluation team will draw on its experience of conducting qualitative and quantitative fieldwork, including with children to ensure that these standards are met, and seek further review where appropriate, to adhere to ethical protocols in line with the OECD-DAC principles of accuracy and credibility.

285. An important consideration, for both adults and children, when seeking their participation in research, is to ensure that they understand exactly what is being done with the information they have provided. The evaluation team will seek to follow a set of ethical principles in conducting all the fieldwork that has been developed based on experience as well as adapted from the Young Live research ethics guidelines,²⁷ which draws from existing literature on the governance of social research²⁸, as well as existing protocols for the protection of children from abuse.²⁹

- **Informed consent:** means that potential respondents are given enough information about the research and researchers ensure that there is no explicit or implicit coercion so that potential respondents can make an informed and free decision on their possible involvement in the fieldwork. This permission will be sought from both the primary caregiver (a teacher if the child is in school) as well children themselves.
- **Anonymity:** given that research respondents share considerable amounts of personal information it is EDOREN's responsibility to ensure that their confidentiality is maintained and personal information is protected for qualitative and quantitative fieldwork. This will be operationalized by ensuring that all datasets are anonymised, in the sense that all names of respondents are removed before the data is shared publically.
- **Ensuring the safety of participants:** this means that the environment in which research is conducted is physically safe. The impact evaluation team will seek to achieve this by ensuring that fieldworkers are local to areas in which they are assigned. In addition fieldwork supervisors will support the fieldwork manager in monitoring local security concerns.

D.2.2.1 Ethical considerations when working with children

286. The following ethical principles are particularly relevant when working with children:

- **Is participation of children genuinely necessary:** is participation strictly necessary to meet the needs of the research enquiry?

²⁷ Young Lives (2011) *Research Ethics*. Young Lives Method Guides

²⁸ Economic and Social Research Council (2010) *Framework for Research Ethics*. ESRC.

²⁹ Save the Children (2003) *Child Protection Policy*. Save the Children Alliance.

- **Recognising that child participants are vulnerable:** the exercise should be carried out with full respect – children for example can easily be treated as inferior and such power dynamics need to be understood and purposefully mitigated in planning and implementation by researchers, enumerators and facilitators.
- **Ensuring that the consent of appropriate adults (parents, head teacher) is sought for a child's participation:** as well as ensuring that the child gives their own informed consent to take part (is happy to do and is not coerced in any way). Children's participation in research must be fully informed and children must also be free not to participate.
- **Ensuring confidentiality, except in circumstances where a child discloses information about a serious child protection concern:** in which case the researchers may need to act (reporting, immediate protection).
- **Ensuring safety of child participants:** If this is in any doubt, at any stage, the activity should be halted.

D.3 DAC Evaluation criteria

287. It is useful to consider the following five DAC criteria when evaluating programmes and projects. These criteria are the parameters based on which the information generated by the evaluation is weighed. The importance given to each of these criteria in the evaluation will depend on the programme being evaluated and its objectives. These criteria are used to structure the evaluation questions above, in line with the proposal in EDOREN's inception report (EDOREN 2013).

D.3.1 Effectiveness

288. This is a measure of the extent to which an aid activity attains its objectives. The following questions may be considered in evaluating the effectiveness of a program of project:

- To what extent were the objectives achieved or are likely to be achieved?
- What were the major factors influencing the achievement or non-achievement of the objectives?

D.3.2 Efficiency

289. Efficiency measures outputs, which could be either quantitative or qualitative, in relation to inputs. This is often measured by comparing different approaches to achieving the same outputs to see whether the most cost-effective process was adopted. It is useful to consider the following questions:

- Were activities cost-efficient?
- Were objectives achieved on time?
- Was the programme or project implemented in the most efficient way compared to alternatives?

D.3.3 Relevance

290. This refers to the extent to which the aid activity is suited to the priorities of the target group, recipient and donor. The following questions may be considered –

- To what extent are objectives of the programme still valid?
- Are the activities and outputs of the programme consistent with the overall goal and the attainment of its objectives?
- Are the activities and outputs of the programme consistent with the intended impacts and effects?

D.3.4 Impact

291. This refers to the positive and negative changes produced by a development intervention, directly or indirectly, intended or unintended. This involves the main impacts resulting from the activity on the local social, economic, environmental and other development indicators. Intended and unintended effects must be taken into account. The following questions may be considered –

- ix. What has happened as a result of the programme or project?
- x. What real difference has the activity made to the beneficiaries?
- xi. How many people have been affected?

D.3.5 Sustainability

292. Sustainability is concerned with measuring whether the benefits of an activity are likely to continue after donor funding has been withdrawn. Projects should be environmentally and financially sustainable. The following questions may be considered -

- xii. To what extent did the benefits of a programme or project continue after donor funding ceased?
- xiii. What were the major factors which influenced the achievement or non-achievement of sustainability of the programme or project?

Annex E Description of Methodology for Calculating the Minimum Detectable Effect (MDE) for a Difference of Differences Estimate

Since a major objective of the evaluation is to measure the difference between the baseline and endline estimates of key indicators for the treatment and control groups, it is ideal to measure the minimum detectable effect size (MDE) for estimates of the "difference of differences". The MDE is defined as the smallest true treatment effect that a research design can detect with confidence. Technically, it is the smallest true treatment effect that has a specified level of statistical power for a particular level of statistical significance, given a specific statistical test. In his paper on "The Core Analytics of Randomized Experiments for Social Research"³⁰, Howard S. Bloom described the MDE as follows (composite from different paragraphs):

The minimum detectable effect of an impact estimator is a multiple of its standard error. Because t-values are multiples of the standard error of an impact estimator, the minimum detectable effect is either $t_{\alpha} + t_{1-\beta}$ (for a one-tail test) or $t_{\alpha/2} + t_{1-\beta}$ (for a two-tail test) times the standard error. These critical t values depend on the number of degrees of freedom. A common convention for defining minimum detectable effects is to set statistical significance (α) at 0.05 and statistical power ($1 - \beta$) at 80 percent. When the number of degrees of freedom exceeds about 20, the multiplier equals roughly 2.5 for a one-tail test and 2.8 for a two-tail test.

One advantage of using a panel survey of schools, teachers and pupils for the evaluation is that the correlation between the baseline and endline samples will increase the level of precision of the estimated differences. The variance (square of the standard error) of the difference between estimates from panel surveys for two different periods can be expressed as follows:

$$\text{var}(\hat{\mu}_2 - \hat{\mu}_1) = (1 - f) \times \left[\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} - \frac{2\rho_{12}\sigma_1\sigma_2}{\sqrt{n_1n_2}} \right] \times deff$$

where:

- μ_2 = estimate of indicator from endline survey
- μ_1 = estimate of indicator from baseline survey
- f = sampling fraction, or proportion of units (schools, teachers or pupils) selected for the baseline and endline surveys
- σ_1 = standard deviation for particular variable in baseline data
- n_1 = sample size for baseline survey
- σ_2 = standard deviation for particular variable in endline data
- n_2 = sample size for endline survey
- ρ_{12} = correlation between values of the variable from the baseline and endline surveys; that is, the inter-temporal correlation (ITC)
- $deff$ = design effect from clustering in the survey data

The variance components inside the brackets of this formula correspond to the variance of the difference between two survey estimates based on simple random sampling (without replacement), and the design effect measures the effect of clustering in the sample design. The design effect is defined as the ratio between the variance of a survey estimate based on the actual clustered design and the corresponding

³⁰ *The Core Analytics of Randomized Experiments for Social Research*, MDRC Working Papers on Research Methodology, Howard S. Bloom, August 2006.

variance based on a simple random sample of the same size. The factor $(1-f)$ is a finite population correction factor when the sampling fraction is high (for example, greater than 0.05). It should be pointed out that in the case of a quasi-experimental design such as the TDP evaluation, where a purposive sample of schools is chosen for the project intervention, the "universe" to which we are making inferences would be the TDP schools in each LGEA, as well as the teachers and pupils in these schools. In the case of the TDP schools, for the baseline and endline surveys we would be selecting 4 out of the 12 TDP schools in the cluster from each LGEA, for a sampling fraction of 0.333. In the case of the "universe" of teacher and pupil indicators it is necessary to determine the total number of teachers and pupils in the TDP schools that are in-scope for the particular indicators (for example, all teachers in the TDP program, or the P2 pupils taught by these teachers. The sampling fraction would correspond to the proportion of these teachers or pupils included in the evaluation survey.

In the expression for the variance of the estimate of a difference shown above, the inter-temporal correlation coefficient (ITC or ρ_{12}) is much higher for a panel of sample schools, teachers, or pupils than for corresponding independent cross-sectional samples for the baseline and endline surveys. It should not be difficult to return to the baseline sample schools for the endline survey, although it is possible that a few sample schools may close or change their grade structure. However, in the case of teachers and pupils, panelling is more challenging and subject to attrition, since teachers and pupils can move or change schools. In order to calculate the MDE, once the sample size is determined for teachers and pupils, it is also necessary to estimate the level of attrition. This will make it possible to estimate the approximate effective sample size for the panel corresponding to the number of baseline teachers and pupils who can be tested for the endline survey.

The design effect (*deff*) mostly measures the effect of clustering in the sample design. The design effect due to clustering is calculated as follows:

$$deff = 1 + \rho_x \times \left(\bar{n} - 1 \right),$$

where:

deff = design effect for the estimate of an indicator based on the TDP sample design

ρ_x = intraclass correlation coefficient (ICC, measure of similarity of schools, teachers or pupils within the cluster, depending on the unit of analysis) for the characteristic being measured

\bar{n} = average number of sampling units (schools, teachers or pupils) per cluster

Since the TDP schools for the baseline survey are selected as a random systematic sample of all the schools in the TPD treatment group, there is no clustering effect for school indicators (that is, the *deff* is equal to 1). In the case of the teacher and pupil indicators, there is a within-school clustering effect. In order to calculate the *deff*, the value of the ICC was estimated based on the results of previous evaluation surveys. In the case of the student test scores, the following ICC estimates were obtained from a paper by Servaas Van Der Berg³¹, based on the average result for several African countries:

ICC for literacy tests = 0.30

³¹ Servaas van der Berg. How effective are poor schools? Poverty and educational outcomes in South Africa. A working paper of the Department of Economics and the Bureau for Economic Research at the University of Stellenbosch.

ICC for numeracy tests = 0.22

In the case of teachers tests, we did not find a good source for the ICC at the time of writing this report, so we assumed a reasonable value of 0.30 for both literacy and numeracy.

In order to calculate the MDE it is also necessary to make an assumption about the inter-temporal correlation (ITC) between the teachers and pupils in the baseline and endline surveys. Given the plan to panel the teachers and pupils as much as possible (that is, the baseline teachers and pupils will be tested again for the endline survey), there should be a high ITC, for example close to 0.8 or 0.9. For the purpose of calculating the MDE, the more conservative value of 0.8 was used for the ITC.

The formula for the variance of the estimate of a difference described above is presented in the general form that can be used for continuous variables as well as proportions. In the case of continuous variables such as test scores, estimates of the mean value and standard deviation of each variable can be obtained from previous education evaluation survey data. The values of the pupil test scores and standard deviations were obtained from the first Composite Survey data; these values were standardized to a scale of 0 to 100 for comparability purposes. These results are presented in Table A1.

Table A1. Estimates of standardized mean pupil test scores and standard deviations from CS1 data

Grade	Subject	Score	SD
P2	Literacy	35.3	30.0
P2	Numeracy	52.1	26.3
P4	Literacy	39.6	31.5
P4	Numeracy	40.3	20.0

In order to obtain values for the mean scores and standard deviations for the teacher tests for the calculation of the MDE, we used the results from the GEP3 TDNA scores for Katsina State, found in a report by David Johnson and Pei-tseng Jenny Hsieh.³² These values are presented below in Table A2.

Table A2. Estimates of standardized mean teacher TDNA scores and standard deviations from GEP3 data for Katsina State

Subject	Score	SD
Teaching primary mathematics	47.5	22.7
Teaching primary English	8.1	20.7

Some of the survey indicators will be in the form of relative estimates such as proportions. In the case of an indicator that is a proportion (p), the standard deviation is equal to $p \times (1-p)$, so we can make a simple calculation of the MDE for a particular proportion estimate using the following formula:

$$\text{var}(\hat{p}_2 - \hat{p}_1) = (1-f) \times \left[\frac{\hat{p}_1 \times (1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2 \times (1-\hat{p}_2)}{n_2} - \frac{2\rho_{12} \sqrt{\hat{p}_1 \times (1-\hat{p}_1) \times \hat{p}_2 \times (1-\hat{p}_2)}}{\sqrt{n_1 n_2}} \right] \times deff,$$

³² David Johnson and Pei-tseng Jenny Hsieh. March 2014. A Diagnostic Assessment of the Professional Development Needs of Teachers in Nigeria, Report Number 4: The characteristics, knowledge and skills of teachers in Katsina State.

where:

$\hat{p}_1 =$ estimate of proportion indicator from baseline survey data

$\hat{p}_2 =$ estimate of proportion indicator from endline survey data

In the case of estimates of proportions, the variance is maximized when the value of the indicator is 0.5 (that is, 50 percent). Therefore we can estimate the maximum standard error using the formula above under the assumption that p_2 has a value of 0.5.

In the case of an estimate of the difference of differences, the variance would be calculated separately for the treatment difference and the control difference, and these variances would be summed, as follows:

$$\text{var}\left(\hat{d}_T - \hat{d}_C\right) = \text{var}\left(\hat{d}_T\right) + \text{var}\left(\hat{d}_C\right),$$

where:

$\text{var}\left(\hat{d}_T\right) =$ variance of difference for treatment group

$\text{var}\left(\hat{d}_C\right) =$ variance of difference for control group

In this case there is no covariance term since the treatment and control samples are considered to be independent. The MDE for the difference of differences can then be estimated as follows:

$$MDES = 2.5 \times se\left(\hat{d}_T - \hat{d}_C\right) = 2.5 \times \sqrt{\text{var}\left(\hat{d}_T\right) + \text{var}\left(\hat{d}_C\right)}$$

Since the alternative hypothesis is that the treatment schools improved more than the control schools, a one-tailed test would be appropriate in this case.

Annex F 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.

6.4 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 5-7 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.

Please ensure that the four ESSPIN school-clusters in an LGA are ***as similar to each other as much 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
- 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 5-7 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 5-7 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!

6.5 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 for 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 S&T training in 2015/2016.

6.6 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 G Evaluation matrix

Reference	Evaluation criterion	Theory of change component	Theory of change detail	Evaluation question	Indicator	Analytical approach	Information collection methods	Disaggregation
Relevance: Does the TDP's in-service training approach and design address needs, priorities and constraints of the primary teachers in northern Nigeria?								
Re-1	Relevance	Final impact	Improved learning English, maths and Science & technology for cohorts taught by selected teachers in TDP schools 2014-2019	Does this objective address needs, priorities and constraints of students in northern Nigeria?	Student learning outcomes	Tabulation from survey data	TDP evaluation quantitative surveys	State, gender
Re-2	Relevance	Intermediate impact	Improved teacher effectiveness in classroom	Does this objective address needs, priorities and constraints of primary teachers in northern Nigeria?	% time teacher involves students in positive interaction in lesson; % days absent from school in year	Tabulation from survey data	TDP evaluation quantitative surveys	State
Re-3	Relevance	Outcome	Improved teacher subject-pedagogical knowledge	Does this objective address needs, priorities and constraints of primary teachers in northern Nigeria?	Teacher TDNA score	Tabulation from survey data	TDP evaluation quantitative surveys	State
Re-4	Relevance	Output	take the lead in developing a model for standardised teacher training, and then coordinating the application of this model across the TDP, ESSPIN and GEP3,	Is this approach coherent with the broader policy environment in at the State and Federal levels in Nigeria, and of DFID?	Is TDP designed appropriately in relation to GEP3 and ESSPIN?	Inference from qualitative data	TDP evaluation qualitative research - ESSPIN, GEP3 interviews; policy documents	State
Re-5	Relevance	Output	develop strategic partnerships with the FMOE, UBEC and NERDC to ensure the national roll out and long-term sustainability of the in-service training model,	Is this approach coherent with the broader policy environment in at the State and Federal levels in Nigeria, and of DFID?	Are FMOE, UBEC and NERDC appropriate custodians of in-service teacher training in Nigeria?	Inference from qualitative data	TDP evaluation qualitative research - FMOE, UBEC, NERDC interviews; policy documents	NA
Re-6	Relevance	Output	partner with SUBEBs that will become key strategic homes for the TDP.	Is this approach coherent with the broader policy environment in at the State and Federal levels in Nigeria, and of DFID?	Are SUBEBs appropriate institutional homes for in-service training in each state?	Inference from qualitative data	TDP evaluation qualitative research - SUBEB, FMOE, UBEC, NERDC interviews; policy documents	State
Re-7	Relevance	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,	Is this approach coherent with the broader policy environment in at the State and Federal levels in Nigeria, and of DFID?	Is a programme of continuous support to teachers realistic and aligned with the policy context in Nigeria?	Inference from qualitative data	TDP evaluation qualitative research - FMOE, UBEC, NERDC interviews; policy documents	State
Re-8	Relevance	Output	engage with the Education Resource Development Council at state level for teacher training material development and strengthen its capacity to become	Is this approach coherent with the broader policy environment in at the State and Federal levels in Nigeria, and of DFID?	EDRC is the appropriate institution for	Inference from qualitative data	TDP evaluation qualitative research - EDRC, SUBEB interviews	State

Reference	Evaluation criterion	Theory of change component	Theory of change detail	Evaluation question	Indicator	Analytical approach	Information collection methods	Disaggregation
			the custodian of all audio-visual resources within and beyond the project period,		material and technology			
Re-9	Relevance	Output	produce Hausa based instructional materials for teachers on maths P1-3,	Does this approach address needs, priorities and constraints of primary teachers in northern Nigeria?	Teachers using Hausa to teach maths 1-3	Inference from qualitative data	TDP evaluation qualitative research - teacher interviews, classroom observations	State
Re-10	Relevance	Output	use English as the language of instruction for all materials from P4-6 and JSS1-3.	Does this approach address needs, priorities and constraints of primary teachers and children in northern Nigeria?	Teachers using English to teach P4-6	Inference from qualitative data	TDP evaluation qualitative research - teacher interviews, classroom observations	State
Re-11	Relevance	Output	select appropriate technology based on an assessment against a range of pre-defined criteria,	Does this approach address needs, priorities and constraints of primary teachers in northern Nigeria?	Teachers being able to use TDP technology	Inference from qualitative data	TDP evaluation qualitative research - teacher interviews, classroom observations	State
Re-12	Relevance	Output	support states to expand teacher support to other schools in the state from 2015 onwards.	Is this approach coherent with the broader policy environment in at the State and Federal levels in Nigeria, and of DFID?	Is the application of the TDP model across TDP states appropriate and consistent with the state policy context?	Inference from qualitative data	TDP evaluation qualitative research - SUBEB interviews; policy documents	State
Relevance: Are TDP's assumptions correct globally and particularly for the Nigerian education and policy context?								
Re-13	Relevance	Intermediate impact to final impact assumption	Children attending school regularly;	Is this assumption correct in the Nigerian context?	% pupils missing more than one day in past month; mean number of days missed; % pupils spending more than 5 hours a day in school	Review of survey estimates	NEDS 2015	State; gender
Re-14	Relevance	Intermediate impact to final impact assumption	Children having the capacity to learn from improved teaching in the language of instruction (they are school ready);	Is this assumption correct in the Nigerian context?	Students reporting understanding materials and curriculum	Inference from qualitative data	TDP evaluation qualitative research - student interviews, classroom observations	State
Re-15	Relevance	Intermediate impact to final impact assumption	Children receiving adequate support for learning at home;	Is this assumption correct in the Nigerian context?	Mean time spent on homework per week; % children receiving	Review of survey estimates	NEDS 2015	Region; gender

Reference	Evaluation criterion	Theory of change component	Theory of change detail	Evaluation question	Indicator	Analytical approach	Information collection methods	Disaggregation
					assistance with homework			
Re-16	Relevance	Intermediate impact to final impact assumption	A class size small enough to allow improved teacher effectiveness to have an impact;	Is this assumption correct in the Nigerian context?	STR at TDP treatment and control schools	Average STR in survey sample	TDP evaluation quantitative surveys	State
Re-17	Relevance	Intermediate impact to final impact assumption	Adequate classroom materials (blackboards, books, desks, etc.) being available	Is this assumption correct in the Nigerian context?	Number of teachers demonstrating adequate classroom materials	Inference from qualitative data	TDP evaluation qualitative research - teacher interviews, classroom observations	State
Re-18	Relevance	Intermediate impact to final impact assumption	Curriculum and materials that are appropriate to the language and ability of students.	Is this assumption correct in the Nigerian context?	Students understanding materials and curriculum	Inference from qualitative data	TDP evaluation qualitative research - student interviews, teacher interviews, classroom observations	State
Re-19	Relevance	Outcome to intermediate impact assumption	TDP materials being appropriate and available	Is this assumption correct in the Nigerian context?	Teachers and government officials reporting TDP materials appropriate, for male and female teachers	Inference from qualitative data	TDP evaluation qualitative research - teacher interviews, LG EA interviews, SUBEB interviews	State; gender
Re-20	Relevance	Outcome to intermediate impact assumption	Selected teachers being retained in schools where the TDP is operating,	Is this assumption correct in the Nigerian context?	% teachers in TDP schools transferred in previous year	Tabulation from survey data	TDP evaluation quantitative surveys	State
Re-21	Relevance	Scale assumption	SUBEBs remain the appropriate institutional home for in-service teacher training	Is this assumption correct in the Nigerian context?	Are SUBEBs appropriate institutional homes for in-service training in each state?	Inference from qualitative data	TDP evaluation qualitative research - SUBEB, FMOE, UBEC, NERDC interviews; policy documents	State
Re-22	Relevance	Scale assumption	State governments/SUBEBs consider in-service teacher training a priority, and the TDP model appropriate, on a sustained basis	Is this assumption correct in the Nigerian context?	Do state governments/SUBEBs consider improving in-service teacher training through the TDP model a priority?	Inference from qualitative data	TDP evaluation qualitative research - SUBEB, FMOE, UBEC, NERDC interviews; policy documents	State
Re-23	Relevance	Scale assumption	SUBEBs have the capacity (managerial, technical and financial) to apply TDP model in other schools	Is this assumption correct in the Nigerian context?	Has TDP adequately designed an approach to account for variations in	Inference from qualitative data	TDP evaluation qualitative research - SUBEB, FMOE, UBEC,	State

Reference	Evaluation criterion	Theory of change component	Theory of change detail	Evaluation question	Indicator	Analytical approach	Information collection methods	Disaggregation
					states' capacity to adopt the TDP?		NERDC interviews; policy documents	
Re-24	Relevance	Sustainability assumption	TDP's partner institutions are the appropriate institutional homes for an in-service teacher training model,	Is this assumption correct in the Nigerian context?	Are FMOE, UBEC and NERDC appropriate custodians of in-service teacher training in Nigeria?	Inference from qualitative data	TDP evaluation qualitative research - FMOE, UBEC, NERDC interviews; policy documents	NA
Re-25	Relevance	Sustainability assumption	reform to teacher effectiveness is an appropriate policy priority to improve learning in Nigeria,	Is this assumption correct in the Nigerian context?	Is teacher effectiveness reform a policy priority in Nigeria?	Inference from qualitative data	TDP evaluation qualitative research - SUBEB, FMOE, UBEC, NERDC interviews; policy documents	NA
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?								
Effi-1	Efficiency	Final impact	Improved learning English, maths and Science & technology for cohorts taught by selected teachers in TDP schools 2014-2019	Does the TDP offer value for money?	TDP expenditure per %age point improvement in learning	Total expenditure on in-service component divided by difference-in-difference of survey data	TDP evaluation quantitative surveys; programme accounts to DFID	State
Effi-2	Efficiency	Intermediate impact	Improved teacher effectiveness in classroom	Does the TDP offer value for money?	TDP expenditure per %age point improvement in time spent on positive interaction in classroom; TDP expenditure per %age point improvement in teacher absenteeism	Total expenditure on in-service component divided by difference-in-difference of survey data	TDP evaluation quantitative surveys; programme accounts to DFID	State
Effi-3	Efficiency	Outcome	Improved teacher subject-pedagogical knowledge	Does the TDP offer value for money?	TDP expenditure per %age point improvement in TDNA scores	Total expenditure on in-service component divided by difference-in-difference of survey data	TDP evaluation quantitative surveys; programme accounts to DFID	State
Effi-4	Efficiency	Output	Implement a two year pilot with 6,000 teachers from 1,500 schools across all LGEAs in three initial states before scaling up to six states,	Does the TDP offer value for money?	TDP expenditure per teacher trained	Total expenditure on in-service component divided by number of teachers trained;	TDP programme monitoring data; programme accounts data to DFID	Fixed and variable costs; state; year

Reference	Evaluation criterion	Theory of change component	Theory of change detail	Evaluation question	Indicator	Analytical approach	Information collection methods	Disaggregation
						comparison with business case		
Effi-5	Efficiency	Output	Implement a two year pilot with 6,000 teachers from 1,500 schools across all LGEAs in three initial states before scaling up to six states,	Were TDP results achieved on time and in full?	Date activity undertaken against plan; reason for deviation	Process evaluation	TDP programme monitoring data	State
Effi-6	Efficiency	Output	Recruit a pool of teacher educators as the teacher development team,	Were TDP results achieved on time and in full?	Date activity undertaken against plan; reason for deviation	Process evaluation	TDP programme monitoring data	State
Effi-7	Efficiency	Output	Recruit teacher facilitators from the current school supervisor cadre,	Were TDP results achieved on time and in full?	Date activity undertaken against plan; reason for deviation	Process evaluation	TDP programme monitoring data	State
Effi-8	Efficiency	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,	Were TDP results achieved on time and in full?	Frequency of inspection visits to schools	Tabulation from survey data	TDP evaluation quantitative surveys	State
Effi-9	Efficiency	Output	Design and implement separate professional development programmes and activities for teachers, head teachers, teacher facilitators and teacher development teams.	Were TDP results achieved on time and in full?	Date activity undertaken against plan; reason for deviation	Process evaluation	TDP programme monitoring data	State
Effi-10	Efficiency	Output	Develop new or adapt existing materials for teachers, teacher facilitators and head teachers,	Were TDP results achieved on time and in full?	Date activity undertaken against plan; reason for deviation	Process evaluation	TDP programme monitoring data	State
Effi-11	Efficiency	Output	Select appropriate technology based on an assessment against a range of pre-defined criteria,	Were TDP results achieved on time and in full?	Date activity undertaken against plan; reason for deviation	Process evaluation	TDP programme monitoring data	State
Effi-12	Efficiency	Output	Produce audio-visual resources on secure digital (SD) cards,	Were TDP results achieved on time and in full?	Date activity undertaken against plan; reason for deviation	Process evaluation	TDP programme monitoring data	State
Effi-13	Efficiency	Output	Engage head teachers actively in pedagogy and leadership and management training,	Were TDP results achieved on time and in full?	Date activity undertaken against plan; reason for deviation	Process evaluation	TDP programme monitoring data	State
Effi-14	Efficiency	Output	Implement a two year pilot with 6,000 teachers from 1,500 schools across all LGEAs in three initial states before scaling up to six states,	How does the TDP's organisational management and set up facilitate delivery?	Efficiency of TDP's organisational management and set up	Process evaluation	TDP evaluation qualitative research - TDP interviews, programme documents	

Reference	Evaluation criterion	Theory of change component	Theory of change detail	Evaluation question	Indicator	Analytical approach	Information collection methods	Disaggregation
Effi-15	Efficiency	Output	Recruit a pool of teacher educators as the teacher development team,	How does the TDP's organisational management and set up facilitate delivery?	Efficiency of TDP's organisational management and set up	Process evaluation	TDP evaluation qualitative research - TDP interviews, programme documents	
Effi-16	Efficiency	Output	Recruit teacher facilitators from the current school supervisor cadre,	How does the TDP's organisational management and set up facilitate delivery?	Efficiency of TDP's organisational management and set up	Process evaluation	TDP evaluation qualitative research - TDP interviews, programme documents	
Effi-17	Efficiency	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,	How does the TDP's organisational management and set up facilitate delivery?	Efficiency of TDP's organisational management and set up	Process evaluation	TDP evaluation qualitative research - TDP interviews, programme documents	
Effi-18	Efficiency	Output	Design and implement separate professional development programmes and activities for teachers, head teachers, teacher facilitators and teacher development teams.	How does the TDP's organisational management and set up facilitate delivery?	Efficiency of TDP's organisational management and set up	Process evaluation	TDP evaluation qualitative research - TDP interviews, programme documents	
Effi-19	Efficiency	Output	Develop new or adapt existing materials for teachers, teacher facilitators and head teachers,	How does the TDP's organisational management and set up facilitate delivery?	Efficiency of TDP's organisational management and set up	Process evaluation	TDP evaluation qualitative research - TDP interviews, programme documents	
Effi-20	Efficiency	Output	Select appropriate technology based on an assessment against a range of pre-defined criteria,	How does the TDP's organisational management and set up facilitate delivery?	Efficiency of TDP's organisational management and set up	Process evaluation	TDP evaluation qualitative research - TDP interviews, programme documents	
Effi-21	Efficiency	Output	Produce audio-visual resources on secure digital (SD) cards,	How does the TDP's organisational management and set up facilitate delivery?	Efficiency of TDP's organisational management and set up	Process evaluation	TDP evaluation qualitative research - TDP interviews, programme documents	
Effi-22	Efficiency	Output	Engage head teachers actively in pedagogy and leadership and management training,	How does the TDP's organisational management and set up facilitate delivery?	Efficiency of TDP's organisational management and set up	Process evaluation	TDP evaluation qualitative research - TDP interviews, programme documents	
Effi-23	Efficiency	Output	Ensure state contributions for procurement and management of technology devices,	What are the contributions of counterparts to TDP's activities?	Counterpart expenditure	Process evaluation	State budget documents, outturn	State, year
Effectiveness: Has TDP led to changes in teacher effectiveness?								
Effe-1	Effectiveness	Intermediate impact	Improved teacher effectiveness in classroom	Has the TDP improved teacher effectiveness in classroom?	% change in proportion of time teacher involves students in positive interaction in a lesson	Constrained randomisation of treatment/control schools; difference-in-difference on	TDP evaluation quantitative surveys - Classroom observations of 1,344 lessons in 2014, 2015, 2017	State

Reference	Evaluation criterion	Theory of change component	Theory of change detail	Evaluation question	Indicator	Analytical approach	Information collection methods	Disaggregation
						treatment and control teachers		
Effe-2	Effectiveness	Intermediate impact	Improved teacher effectiveness in classroom	Has the TDP improved teacher effectiveness in classroom?	% change in % of working days for which teacher absent in last year; reasons for teacher absenteeism	Constrained randomisation of treatment/control schools; difference-in-difference on treatment and control teachers	TDP evaluation quantitative surveys - Self-reported absenteeism cross checked with school records for 1,344 teachers in 336 schools	State
Effe-3	Effectiveness	Intermediate impact	Improved teacher effectiveness in outside classroom support	Has the TDP improved teacher effectiveness in outside classroom support?	% change in number of teachers testing pupils recently	Constrained randomisation of treatment/control schools; difference-in-difference on treatment and control teachers	TDP evaluation quantitative surveys - Self-reported testing for 1,344 teachers in 336 schools	State
Effe-4	Effectiveness	Outcome	Improved teacher subject-pedagogical knowledge	Has the TDP improved teacher knowledge?	% change in average teacher raw scores on TDNA for English	Constrained randomisation of treatment/control schools; difference-in-difference on treatment and control teachers	TDP evaluation quantitative surveys - TDNA of 1,344 teachers	State, gender of teacher
Effe-5	Effectiveness	Outcome	Improved head teacher leadership and management	Has the TDP improved head teacher leadership and management?	% change in number of head teacher meetings with teachers; % change in number of head teacher lesson observations; % change in number of head teachers taking action to reduce absenteeism	Constrained randomisation of treatment/control schools; difference-in-difference on treatment and control teachers	TDP evaluation quantitative surveys - head teacher interviews in 336 schools	
Effe-6	Effectiveness	Outcome	Improved teacher pedagogical knowledge	Has the TDP improved teacher pedagogical knowledge	% change in teacher pedagogical scores	Constrained randomisation of treatment/control schools; difference-in-difference on	TDP evaluation quantitative surveys - TDNA of 1,344 teachers	State, gender of teacher

Reference	Evaluation criterion	Theory of change component	Theory of change detail	Evaluation question	Indicator	Analytical approach	Information collection methods	Disaggregation
						treatment and control teachers		
Effe-7	Effectiveness	Output	Implement a two year pilot with 6,000 teachers from 1,500 schools across all LGEAs in three initial states before scaling up to six states,	Has the TDP trained more teachers?	% change in number of teachers receiving in-service training; regularity of training attendance	Constrained randomisation of treatment/control schools; difference-in-difference on treatment and control teachers	TDP evaluation quantitative surveys - TDNA of 1,344 teachers	State
Effe-8	Effectiveness	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,	Has the TDP improved support to schools?	% change in the frequency of inspection visits to schools	Constrained randomisation of treatment/control schools; difference-in-difference on treatment and control teachers	TDP evaluation quantitative surveys - head teacher interviews in 336 schools	State
Effe-9	Effectiveness	Output	Develop new or adapt existing materials for teachers, teacher facilitators and head teachers,	Has the TDP improved teacher materials?	Perception of usefulness of materials	Inference from qualitative data	TDP evaluation qualitative research - student interviews, teacher interviews, classroom observations	State
Effe-10	Effectiveness	Output	Develop new or adapt existing materials for teachers, teacher facilitators and head teachers,	Has the TDP improved teacher materials?	% teachers reporting use of TDP materials	Constrained randomisation of treatment/control schools; difference-in-difference on treatment and control teachers	TDP evaluation quantitative surveys - 1,344 teacher interviews	State
Effe-11	Effectiveness	Output	Engage head teachers actively in pedagogy and leadership and management training,	Has the TDP trained head teachers?	% change in number of teachers receiving in-service training; regularity of training attendance	Constrained randomisation of treatment/control schools; difference-in-difference on treatment and control teachers	TDP evaluation quantitative surveys - head teacher interviews in 336 schools	State
Effe-12	Effectiveness	Outcome to intermediate impact assumption	TDP materials being appropriate and available	What factors facilitated or inhibited the TDP achieving its outcomes?	Teachers and government officials reporting TDP materials appropriate	Inference from qualitative data	TDP evaluation qualitative research - teacher interviews, LGEA	State; gender

Reference	Evaluation criterion	Theory of change component	Theory of change detail	Evaluation question	Indicator	Analytical approach	Information collection methods	Disaggregation
					and available, for male and female teachers		interviews, SUBEB interviews	
Effe-13	Effectiveness	Outcome to intermediate impact assumption	Selected teachers being sufficiently intrinsically motivated to turn improved knowledge into improved effectiveness,	What factors facilitated or inhibited the TDP achieving its outcomes?	Teachers reporting/displaying adequate intrinsic motivation	Inference from qualitative data	TDP evaluation qualitative research - teacher interviews, classroom observations	State, gender
Effe-14	Effectiveness	Outcome to intermediate impact assumption	Selected teachers being sufficiently extrinsically motivated to apply their new knowledge,	What factors facilitated or inhibited the TDP achieving its outcomes?	% teachers with salaries paid on time and in full; Teachers reporting adequate extrinsic motivation	Tabulation from survey data; inference from qualitative data, secondary data	TDP evaluation survey - 1,344 teacher interviews; TDP evaluation qualitative research - teacher interviews; EDOREN teacher management study	State, gender
Effe-15	Effectiveness	Outcome to intermediate impact assumption	Selected teachers being retained in schools where the TDP is operating,	What factors facilitated or inhibited the TDP achieving its outcomes?	% teachers in TDP schools transferred in previous year	Tabulation from survey data; inference from qualitative data	TDP evaluation quantitative surveys; TDP evaluation qualitative research - head teacher interviews	State
Effe-16	Effectiveness	Outcome to intermediate impact assumption	Selected teachers being class ready, in other words have the capacity to apply their new knowledge,	What factors facilitated or inhibited the TDP achieving its outcomes?	% teachers with appropriate qualifications; % teachers receiving training in last three years	Tabulation from survey data; inference from qualitative data	TDP evaluation quantitative surveys; TDP evaluation qualitative research - head teacher interviews, teacher interviews	State
Effe-17	Effectiveness	Outcome to intermediate impact assumption	Selected teachers being supported to apply their new knowledge	What factors facilitated or inhibited the TDP achieving its outcomes?	Number of head teacher meetings with teachers; number of head teacher lesson observations; number of head teachers taking action to reduce absenteeism	Tabulation from survey data; inference from qualitative data	TDP evaluation quantitative surveys - head teacher interviews in 336 schools; TDP evaluation qualitative research - head teacher, teacher interviews	State

Reference	Evaluation criterion	Theory of change component	Theory of change detail	Evaluation question	Indicator	Analytical approach	Information collection methods	Disaggregation
Effe-18	Effectiveness	Outcome to intermediate impact assumption	Head teachers being motivated to lead and manage teachers well.	What factors facilitated or inhibited the TDP achieving its outcomes?	Number of head teacher meetings with teachers; number of head teacher lesson observations; number of head teachers taking action to reduce absenteeism	Tabulation from survey data; inference from qualitative data	TDP evaluation quantitative surveys - head teacher interviews in 336 schools; TDP evaluation qualitative research - head teacher interviews	State
Effe-19	Effectiveness	Output to outcome assumption	Teacher educators and teacher facilitators are able to absorb and then transfer the skills offered in the TDP training;	What factors facilitated or inhibited the TDP achieving its outcomes?	Teacher educator and teacher facilitator perception of capacity to absorb training	Inference from qualitative data	TDP evaluation qualitative research - teacher facilitator interviews, training observations, teacher interviews	State
Effe-20	Effectiveness	Output to outcome assumption	Teacher educators and facilitators have sufficient time and material support to provide support to schools (i.e. they are not directed to other tasks, transferred, or made redundant);	What factors facilitated or inhibited the TDP achieving its outcomes?	Teacher educator and teacher facilitator perception of adequate time and support	Inference from qualitative data	TDP evaluation qualitative research - teacher facilitator interviews, LGEA interviews, SUBEB interviews	State
Effe-21	Effectiveness	Output to outcome assumption	Head teachers have appropriate incentives/motivation to apply their new knowledge in support of teachers;	What factors facilitated or inhibited the TDP achieving its outcomes?	Head teacher perceptions of incentives/motivation	Inference from qualitative data	TDP evaluation qualitative research - head teacher interviews, teacher interviews	State
Effe-22	Effectiveness	Output to outcome assumption	Teachers and head teachers attend training regularly (i.e. they have the time and are managed and supported to so, and not transferred during the training process);	What factors facilitated or inhibited the TDP achieving its outcomes?	% teachers and head teachers receiving TDP training; frequency of attendance	Tabulation from survey data; inference from qualitative data	TDP evaluation quantitative surveys; TDP evaluation qualitative research - head teacher interviews, teacher interviews	State
Effe-23	Effectiveness	Output to outcome assumption	Teachers can access and use the audio-visual materials (i.e. the technology works, can be charged, is not lost, stolen or broken, is upgraded or fixed where appropriate, can be understood, etc.);	What factors facilitated or inhibited the TDP achieving its outcomes?	% teachers with Smartphone; % teachers with electricity at home; % teachers reporting using 'trainer-in-pocket'; % schools	Tabulation from survey data; inference from qualitative data	TDP evaluation quantitative surveys; TDP evaluation qualitative research - head teacher interviews, teacher interviews	State

Reference	Evaluation criterion	Theory of change component	Theory of change detail	Evaluation question	Indicator	Analytical approach	Information collection methods	Disaggregation
					with facilities for technology			
Effe-24	Effectiveness	Output to outcome assumption	2.5 days' training, monthly cluster meetings, and the support visits and mentoring are sufficient to instil new pedagogical knowledge in teachers;	What factors facilitated or inhibited the TDP achieving its outcomes?	Teacher facilitator, head teacher and teacher perceptions of adequacy of training	Inference from qualitative data	TDP evaluation qualitative research - teacher facilitator interviews, training observations, teacher interviews, head teacher interviews	State
Effe-25	Effectiveness	Output to outcome assumption	Teachers have the basic language, subject and pedagogical skills to absorb the new knowledge and skills available from TDP;	What factors facilitated or inhibited the TDP achieving its outcomes?	Average teacher raw scores on TDNA for English	Tabulation from survey data; inference from qualitative data	TDP evaluation quantitative surveys - TDNA of 1,344 teachers; TDP qualitative research - teacher interviews	State, gender of teacher
Effe-26	Effectiveness	Output to outcome assumption	The materials are appropriate for students of different abilities, particularly around language	What factors facilitated or inhibited the TDP achieving its outcomes?	Teacher and student perceptions of adequacy of materials	Inference from qualitative data	TDP evaluation qualitative research - teacher interviews, classroom observations, student interviews	State
Effe-27	Effectiveness	Output to outcome assumption	There are no changes to the curriculum or other features of the education system that render these materials redundant.	What factors facilitated or inhibited the TDP achieving its outcomes?	Exogenous changes	Inference from qualitative data	TDP evaluation qualitative research - SUBEB interviews; policy documents	State
Effe-28	Effectiveness	Output		What were the synergies between outputs?	NA	Inference from qualitative data	All	State
Impact: Has TDP caused changes in student learning in English, maths and science & technology in target schools?								
Im-1	Impact	Final impact	Improved learning English, maths and Science & technology for cohorts taught by selected teachers in TDP schools 2014-2019	Has the TDP improved learning outcomes for children taught by TDP-trained teachers?	% change in average raw scores and expected competency achievement for students taught by TDP-trained teachers in English, maths and science&technology	Constrained randomisation of treatment/control schools; difference-in-difference on treatment and control cohorts moving from p3 to p6	TDP evaluation quantitative surveys - MLA survey of 2,688 students in 2014 and 2017	State, gender of pupil

Reference	Evaluation criterion	Theory of change component	Theory of change detail	Evaluation question	Indicator	Analytical approach	Information collection methods	Disaggregation
Im-2	Impact	Final impact	Improved learning English, maths and Science & technology for cohorts taught by selected teachers in TDP schools 2014-2019	Are TDP impacts heterogenous amongst teachers with different qualifications and years of experience?	% change in learning outcomes for students	Constrained randomisation of treatment/control schools; difference-in-difference on treatment and control cohorts moving from p3 to p6; regression analysis by teacher qualifications and experience	TDP evaluation quantitative surveys - MLA survey of 2,688 students in 2014 and 2017; 1,344 teacher interviews	Teacher qualifications, years of experience
Im-3	Impact	Final impact		Are there any other positive or negative unanticipated TDP impacts?	NA	Inference from qualitative data; potential test in endline	All	TBD
Im-4	Impact	Final impact		Are there any impacts resulting from synergies between the two TDP components?	NA	Inference from qualitative data; potential test in endline	All	TBD
Im-5	Impact	Scale impact	TDP in-service training improves performance of other teachers in TDP schools	Are there any indications that the TDP is improving teacher performance of teachers not trained by the TDP in TDP schools?	Perceptions of other teacher effectiveness; self-reported application of model	Inference from qualitative data	TDP evaluation qualitative research - head teacher interviews, other teacher interviews, other teacher classroom observations, student interviews	State
Im-6	Impact	Scale impact	TDP model applied in other schools in TDP states	Are there any indications that the TDP model is applied in other schools in TDP states?	Perceptions of application of model in other schools in TDP states	Inference from qualitative data	TDP evaluation qualitative research - SUBEB interviews	State
Im-7	Impact	Scale impact	TDP model applied in other states in Nigeria	Are there any indications that the TDP model is applied in other states?	Perceptions of application of model in other states	Inference from qualitative data	TDP evaluation qualitative research - UBEC, FMOE interviews	State
Im-8	Impact	Intermediate impact to final impact assumption	Children attending school regularly;	What factors facilitated or inhibited the TDP achieving its impacts?	Perceptions of regular attendance by students and teachers; % days attended	Inference from qualitative data	TDP evaluation qualitative research - teacher interviews, student interviews, attendance registers	State

Reference	Evaluation criterion	Theory of change component	Theory of change detail	Evaluation question	Indicator	Analytical approach	Information collection methods	Disaggregation
Im-9	Impact	Intermediate impact to final impact assumption	Children having the capacity to learn from improved teaching in the language of instruction (they are school ready);	What factors facilitated or inhibited the TDP achieving its impacts?	Average student competencies in English; Perceptions of student ability to learn in language of instruction	Tabulation from survey data; inference from qualitative data	TDP evaluation quantitative surveys - MLA survey of 2,688 students in 2014 and 2017; TDP qualitative research - teacher interviews, student interviews, classroom observations	State, gender of student
Im-10	Impact	Intermediate impact to final impact assumption	Children receiving adequate support for learning at home;	What factors facilitated or inhibited the TDP achieving its impacts?	Perceptions of adequate support at home	Inference from qualitative data and NEDS 2015	TDP evaluation qualitative research - teacher interviews, student interviews; NEDS 2015	State, gender of student
Im-11	Impact	Intermediate impact to final impact assumption	Children supporting their peers to learn;	What factors facilitated or inhibited the TDP achieving its impacts?	Perceptions of peer support	Inference from qualitative data	TDP evaluation qualitative research - teacher interviews, student interviews, classroom observations	State
Im-12	Impact	Intermediate impact to final impact assumption	A class size small enough to allow improved teacher effectiveness to have an impact;	What factors facilitated or inhibited the TDP achieving its impacts?	STR; perceptions of adequacy of class size	Inference from qualitative data	TDP qualitative research - teacher interviews, student interviews, classroom observations, school records	State
Im-13	Impact	Intermediate impact to final impact assumption	Adequate classroom materials (blackboards, books, desks, etc.) being available	What factors facilitated or inhibited the TDP achieving its impacts?	Perceptions of adequacy of classroom materials	Inference from qualitative data	TDP qualitative research - teacher interviews, student interviews, classroom observations	State
Im-14	Impact	Intermediate impact to final impact assumption	Curriculum and materials that are appropriate to the language and ability of students.	What factors facilitated or inhibited the TDP achieving its impacts?	Perceptions of students understanding materials and curriculum	Inference from qualitative data	TDP evaluation qualitative research - student interviews, teacher interviews, classroom observations	State

Reference	Evaluation criterion	Theory of change component	Theory of change detail	Evaluation question	Indicator	Analytical approach	Information collection methods	Disaggregation
Im-15	Impact	Scale assumption	Head teachers apply TDP approaches to other teachers, including meetings, support, and monitoring	What factors facilitated or inhibited the TDP achieving its impacts?	Number of head teacher meetings with other teachers; number of head teacher lesson observations of other teachers; number of head teachers taking action to reduce absenteeism; perceptions of wider application by head teachers	Tabulation from survey data; inference from qualitative data	TDP evaluation quantitative surveys - head teacher interviews in 336 schools; TDP evaluation qualitative research - head teacher, non-TDP trained teacher interviews	State
Im-16	Impact	Scale assumption	Other teachers are capable, motivated and managed to adopt TDP approaches with more limited support	What factors facilitated or inhibited the TDP achieving its impacts?	Perceptions of other teacher capacity	Inference from qualitative data	TDP evaluation qualitative research - head teacher interviews, other teacher interviews, other teacher classroom observations, student interviews	State
Im-17	Impact	Scale assumption	TDP materials are shared with other teachers	What factors facilitated or inhibited the TDP achieving its impacts?	Reported sharing of information by head teachers	Inference from qualitative data	TDP evaluation qualitative research - head teacher interviews, other teacher interviews	State
Im-18	Impact	Scale assumption	ESSPIN and GEP3 have the interest and capacity (managerial, technical and financial) to apply TDP model	What factors facilitated or inhibited the TDP achieving its impacts?	Perception of ESSPIN and GEP3 capacity to apply model	Inference from qualitative data	TDP evaluation qualitative research - TDP interviews, ESSPIN interviews, GEP3 interviews, DFID interviews	
Im-19	Impact	Scale assumption	ESSPIN and GEP3 can apply TDP model successfully in their timeframe	What factors facilitated or inhibited the TDP achieving its impacts?	Perception of ESSPIN and GEP3 capacity to apply model	Inference from qualitative data	TDP evaluation qualitative research - TDP interviews, ESSPIN interviews, GEP3 interviews, DFID interviews	
Im-20	Impact	Scale assumption	SUBEBS remain the appropriate institutional home for in-service teacher training	What factors facilitated or inhibited the TDP achieving its impacts?	Perception of SUBEBS as appropriate institutional homes for in-service training in each state	Inference from qualitative data	TDP evaluation qualitative research - SUBEB, FMOE, UBEC, NERDC interviews; policy documents	State
Im-21	Impact	Scale assumption	State governments/SUBEBS consider in-service teacher training a priority, and the TDP model appropriate, on a sustained basis	What factors facilitated or inhibited the TDP achieving its impacts?	Do state governments/SUBEBS consider improving in-service teacher	Inference from qualitative data	TDP evaluation qualitative research - SUBEB, FMOE, UBEC,	State

Reference	Evaluation criterion	Theory of change component	Theory of change detail	Evaluation question	Indicator	Analytical approach	Information collection methods	Disaggregation
					training through the TDP model a priority?		NERDC interviews; policy documents	
Im-22	Impact	Scale assumption	SUBEBS have the capacity (managerial, technical and financial) to apply TDP model in other schools	What factors facilitated or inhibited the TDP achieving its impacts?	Perception of SUBEB capacity to apply TDP model	Inference from qualitative data	TDP evaluation qualitative research - SUBEB interviews, TDP interviews	State
Sustainability: Are TDP's impacts on teacher effectiveness sustainable without further DFID support?								
Su-23	Sustainability	Sustainability impact	TDP model applied sustainably in TDP schools, in other schools in TDP states, and in Nigeria	Are the TDP's impacts likely to be sustainable when DFID ends funding in 2019?	Perceptions of maintenance of TDP approach and practice	Inference from qualitative data	TDP evaluation qualitative research - UBEC interviews, FMOE interviews, SUBEB interviews, LGEA interviews, teacher facilitator interviews, head teacher interviews, teacher interviews	State
Su-24	Sustainability	Sustainability assumption	TDP's partner institutions are the appropriate institutional homes for an in-service teacher training model,	Are the TDP's impacts likely to be sustainable when DFID ends funding in 2019?	FMOE, UBEC and NERDC are appropriate custodians of in-service teacher training in Nigeria	Inference from qualitative data	TDP evaluation qualitative research - FMOE, UBEC, NERDC interviews; policy documents	NA
Su-25	Sustainability	Sustainability assumption	TDP's partner institutions are open to these partnerships being developed, on a continuous basis (e.g. in a way that is resilient to changes in their leadership and political standing),	Are the TDP's impacts likely to be sustainable when DFID ends funding in 2019?	Perception of institutional capacity to apply TDP model	Inference from qualitative data	TDP evaluation qualitative research - SUBEB, FMOE, UBEC, NERDC interviews; TDP interviews, DFID interviews	
Su-26	Sustainability	Sustainability assumption	TDP's partner institutions have the capacity (broadly defined to include e.g. funding, human resources, management capacity, skills, etc.) to engage in the in-service training model,	Are the TDP's impacts likely to be sustainable when DFID ends funding in 2019?	Perception of institutional capacity to apply TDP model	Inference from qualitative data	TDP evaluation qualitative research - SUBEB, FMOE, UBEC, NERDC interviews; TDP interviews, DFID interviews	
Su-27	Sustainability	Sustainability assumption	Reform to teacher effectiveness is an appropriate policy priority to improve learning in Nigeria,	Are the TDP's impacts likely to be sustainable when DFID ends funding in 2019?	Teacher effectiveness reform is a policy priority in Nigeria	Inference from qualitative data	TDP evaluation qualitative research - SUBEB, FMOE, UBEC, NERDC interviews; policy documents	
Su-28	Sustainability	Sustainability assumption	Reform to in-service teacher training is an appropriate policy priority to improve teacher effectiveness in Nigeria,	Are the TDP's impacts likely to be sustainable when DFID ends funding in 2019?	In-service teacher training reform is a priority for teacher effectiveness reform in Nigeria	Inference from qualitative data	TDP evaluation qualitative research - SUBEB, FMOE, UBEC, NERDC interviews; policy documents	NA

Reference	Evaluation criterion	Theory of change component	Theory of change detail	Evaluation question	Indicator	Analytical approach	Information collection methods	Disaggregation
Su-29	Sustainability	Sustainability assumption	A single in-service model is appropriate for all different states in Nigeria (rather than different models in different states),	Are the TDP's impacts likely to be sustainable when DFID ends funding in 2019?	Perception of appropriateness of model	Inference from qualitative data	TDP evaluation qualitative research - SUBEB, FMOE, UBEC, NERDC interviews, TDP interviews, DFID interviews	
Su-30	Sustainability	Sustainability assumption	The TDP in-service model is appropriate in Nigeria, and improves teacher effectiveness and learning,	Are the TDP's impacts likely to be sustainable when DFID ends funding in 2019?	% change in learning outcomes for students	Constrained randomisation of treatment/control schools; difference-in-difference on treatment and control cohorts moving from p3 to p6	TDP evaluation quantitative surveys - MLA survey of 2,688 students in 2014 and 2017	NA
Su-31	Sustainability	Sustainability assumption	Training materials can be updated to reflect changes to the curriculum, language of instruction, etc.,	Are the TDP's impacts likely to be sustainable when DFID ends funding in 2019?	Flexibility of training materials	Inference from qualitative data	TDP evaluation qualitative research - TDP interviews	State, gender of pupil
Su-32	Sustainability	Sustainability assumption	The technology remains robust to technological and infrastructural change and can be maintained in working order.	Are the TDP's impacts likely to be sustainable when DFID ends funding in 2019?	Perception of longevity of technology	Inference from qualitative data	TDP evaluation qualitative research - head teacher interviews, teacher interviews	
Su-33	Sustainability	Sustainability assumption	The SUBEBs and LGEAs have the incentives and capacity (managerial, financial, technical) to maintain, support and renew the teacher educator teams without TDP support, and	Are the TDP's impacts likely to be sustainable when DFID ends funding in 2019?	Perceptions of SUBEB and LGEA capacity to maintain TDP approach and practice	Inference from qualitative data	TDP evaluation qualitative research - UBEC interviews, FMOE interviews, SUBEB interviews, LGEA interviews	State
Su-34	Sustainability	Sustainability assumption	Head teachers and teachers have the interest and capacity to continue to support the TDP in-service training model.	Are the TDP's impacts likely to be sustainable when DFID ends funding in 2019?	Perceptions of head teacher and teacher interest and capacity maintain TDP approach and practice	Inference from qualitative data	TDP evaluation qualitative research - head teacher interviews, teacher interviews	State