

Evaluation learning review for DFID Nigeria

Laura Bolton

IDS

2 April 2019

Question

- 1) How, where and when/under what conditions has Head Teachers and Teachers' capacity building approach through selected programmes led to improved learning outcomes for children? Include primary, with a focus on the early grades.
- 2) What factors were associated with programmes contributing to progress (or lack of progress) in teacher performance or effectiveness, including teachers' motivation, pedagogical skills and subject knowledge?
- 3) What evidence is there of the relative value and effectiveness of different types of teaching and learning materials for the intended audiences (print, audio visual)?
- 4) What evidence is there of the value or merit of the teacher development systems which were adopted by programmes, particularly the use of "state school improvement teams" or "master trainers" or "teacher development teams" or "school support officers" to provide training, coaching and mentoring?
- 5) What factors were associated with DFID programmes contributing to changes or improvements in management systems at state and local government levels (including unintended consequences – positive or negative)?
- 6) What evidence is there of DFID programmes changing and adapting in response to available evidence?
- 7) What has (or has not) worked in terms of gender sensitive programming amongst the programmes surveyed? What has (or has not) worked to enable learning for girls?
- 8) What is the range of different approaches and instruments that have been used to measure teachers' knowledge, teachers' effectiveness and children's learning? What reflections are given on the relative merits of the different approaches?

The K4D helpdesk service provides brief summaries of current research, evidence, and lessons learned. Helpdesk reports are not rigorous or systematic reviews; they are intended to provide an introduction to the most important evidence related to a research question. They draw on a rapid desk-based review of published literature and consultation with subject specialists.

Helpdesk reports are commissioned by the UK Department for International Development and other Government departments, but the views and opinions expressed do not necessarily reflect those of DFID, the UK Government, K4D or any other contributing organisation. For further information, please contact helpdesk@k4d.info.

Contents

1. Summary
2. The evaluations
3. Teachers capacity leading to improved learning outcomes
4. Factors associated with teacher performance or effectiveness
5. Relative value and effectiveness of different types of teaching and learning materials
6. Teacher development systems
7. Changes in management at state or local government level
8. Evidence leading to programme adaptation
9. Gender sensitive programming
10. Approaches to and instruments for measurement
11. References
12. Appendix

1. Summary

Eight evaluations and one research report were provided by DFID for this review and are described in section 2. These were drawn on to answer the eight questions listed above.

Key points on each question:

Teachers capacity leading to improved learning outcomes

Some of the projects did not produce improved learning outcomes. One programme found that coaching alongside teacher training did not improve learning outcomes. Other evaluations were not clear on what conditions had improved results. There are other factors to consider that limit learning outcomes such as absenteeism and time on task.

Factors associated with teacher performance or effectiveness

Training effectiveness may be limited by low initial subject knowledge and pedagogical knowledge; language barriers; and the challenging class environment. Teacher motivation is also dependent on a wide range of factors such as excessive workload, working conditions and salary.

Relative value and effectiveness of different types of teaching and learning materials

Merits of different training materials were largely not discussed within these evaluations.

Teacher development systems

The evaluations reviewed for this report did not provide clear evidence of the merits of different teacher development systems, a small amount of experience was described. The GEP III evaluations made some suggestions to ensure success of peer-mentoring: teachers being literate in Hausa, weekly meetings being feasible given context, lead teachers having capacity, motivation and availability, and Master Trainers having resources, competencies, and incentives. A cascade training model in the teacher development model was given mixed reviews depending on state. The factors affecting positive or negative experience were not described.

Changes in management at state or local government level

Some government capacity development seems to be implicit in GEP III but measurement is not reported on explicitly. Government stakeholders report appreciation for the capacity building opportunities, particularly in the areas of Information Systems and staff working on the Annual School Census.

Programme adaptations made in response to evidence

The evaluations tended to make recommendations for adaptations but did not report on evidence of these changes being implemented. The reports focussed on evidence generated rather than evidence previously used. Within the GEP III programme, the teaching approach was adapted as it was found to be unsuitable. The need to adapt learning materials was highlighted as a priority. The remedial learning programme based on Teaching at the Right Level is noted to be built on a citizen-led household assessment but the specific changes this evidence led to were not described. The teaching approach in RANA was adapted but seemed to be based on a general consensus rather than specific evidence.

Gender-sensitive programming

The GEP III evaluation concluded that it was unlikely that gender-sensitive class practice had improved. This was not an explicit part of the programme implementation and design. In particular, training for facilitators should include a focus on this. The ENGINE II assessment

identifies the importance of streaming for 'out of school' girls to effectively learn. Research on the impacts of female teachers found their presence to be positively correlated with girls' enrolment and the learning of both boys and girls. Other programmes did not have an explicit focus on gender-sensitivity in their goals. The GEP III evaluation notes the concept as difficult to define and reliably measure. Specific success factors in enabling girls to learn were not identified.

Approaches to measurement

GEP III used a highly developed system for assessing teacher knowledge. There was also a highly involved approach to measuring student learning. The merits of these were not discussed. The Teacher Development Programme used interviews and noted the problem of self-reporting in recording teacher motivation and activity. Teacher observation was also cautioned as teachers could adjust their practice as they know they are being observed. Some of the evaluations gave little reflection on the merits of data collection instruments. Early Grade Reading Assessment (EGRA) and Early Grade Mathematics Assessment (EGMA) were commonly used. The Northern Education Initiative plus conducted workshops and piloted different EGRA instruments to develop the best version.

2. The evaluations

The references in this report use the below source codes (for example S1) which match onto the specific quotes detailed in the accompanying appendix (for example S1.1). Please refer to the appendix for information on specific reference page numbers and exact quotations. The quotations in the appendix are provided for further details on each point made if required.

2.1 Girls Education Project Phase 3 (GEP III)

Source 1 (S1): **Midline Evaluation of UNICEF's DFID-funded Girls' Education Project Phase 3**. Pellens et al. (2018a) EDOREN.

Source 2 (S2): **Midline Evaluation of UNICEF's DFID-funded Girls' Education Project Phase 3 Technical Report**, prepared by EDOREN on behalf of UNICEF GEP3. Pellens et al. (2018b). EDOREN.

The Girls Education Project (GEP3) consists of multiple interventions aiming to provide meaningful and relevant quality basic education for integrated Qur'anic education. Increasing capacity of teachers is one output. Other outputs are increasing access and demand and improved governance. The effectiveness evaluations focus on two elements: 1) RANA early learning programme, in-service professional development; and 2) GEP3's Support to integrated Qur'anic schools (IQSs) which included teacher training and mentoring alongside a package of learning materials, and management committee support.

2.2 Teacher Development Programme

Source 3 (S3): **Teacher Development Programme. Endline Evaluation (Volume I)**. Cameron et al., (2018). EDOREN.

And source 3s (S3a). Volume II.

The Teacher Development Programme (TDP) is a six-year (2013–19) programme in Nigeria working through in-service training for primary teachers, reform of pre-service teacher education, and strengthening evidence-based research on teaching. The evaluation focuses on in-service training at primary level in English, mathematics and science.

2.3 Non-Formal Learning Centres (NFLCs)

Source 4 (S4): **The cost-effectiveness of ALP for out-of-school children and the added value and effect of providing on-site coaching to learning facilitators in NFLC on students' learning and transition outcomes in Nigeria: Endline Report.** Diazgranados et al. (2018)

The aim of this programme was to provide 20,000 out-of-school children with access to safe learning through 400 NFLCs. 400 community members were trained in the states of Yobe and Borno and given ongoing professional development opportunities. They went on to teach basic numeracy, literacy, and socio-emotional learning skills to children.

2.4 Reading and Numeracy Activity (RANA)

Source 5 (S5): **Nigeria Reading and Numeracy Activity: End of year report 2017-18.** DFID, UNICEF, FHi360 (2018).

This evaluation looks at the third year of programming in Katsina and Zamfara. RANA provided training to 40 Master Trainers who then provided training to 800 educators. Teachers were further supported by the RANA community of practice. There was a focus on training to teach the syllabic rather than letter sound approach to reading instruction. In addition to practical classroom management techniques to promote pupil participation. The evaluation was not run against a counterfactual.

2.5 Northern Education Initiative Plus

Source 6 (S6): **Northern Education Initiative Plus. Early Grade Reading Assessment Midline Report.** Creative Associates International. (2018). USAID.

This activity, run in Bauchi and Sokoto, aims to improve the states' ability to provide quality education and improve reading skills for more than 2 million school-aged children and youth. Around 45,000 teachers and facilitators are being trained and equipped to reach children in schools and in NFLCs.

2.6 UNICEF Girls Education Project Phase 3. The effect of female teachers on girls enrolment

Source 7 (S7): **What is the effect of female teachers on girls' enrollment and retention in Northern Nigeria? A Mixed-Methods Research Study under the UNICEF/Nigeria Girls'**

Education Project, Phase 3. Smiley, A., Moussa, W., and Brown, E.B. FHI 360 Education Policy and Data Center.

Mixed-methods evidence on the effects of female teachers in schools. Not focussing on one particular intervention.

2.7 Teaching at the Right Level (TaRL)

Source 8 (S8): **Teaching at the Right Level (TaRL). Pilot Programme.** December 2018 Report. The Education Partnership Centre.

A simple approach, which assesses children's reading and mathematics skills and then groups them according to level.

2.8 ENGINE II Learning needs assessment

Source 9 (S9): **Learning Needs Assessment.** Preliminary findings.

This document does not give detail on programme design or results.

2.9 Research on Female Teachers

Smiley, A., Moussa, W., and Brown, E.B. (undated). FHI360, UNICEF, UKAID.

A report on research undertaken within the UNICEF/GEP project phase 3. It looks at the extent to which the presence of female teachers influences girls' enrolment and whether female teachers improve outcomes. It also identifies contributions to the underrepresentation of qualified female teachers.

3. Teachers capacity leading to improved learning outcomes

3.1 GEP III

The RANA programme had a very small significant impact on learning and there was no improvement in the performance gap between highest and lowest performing pupils (S1.1.).

Those at the lowest literacy level (pre-literacy) decreased from 97% at baseline to 90% at midline (S1.2). Literacy levels classed as emerging went from 1% to 6%. Those at the highest literacy level (basic literacy) went from 2% to 4%.

The analysis in this survey finds very few factors significantly affecting learning (S1.3). There was limited variability between base and midline in the factors of interest including teacher competency (S2.1).

3.2 Teacher Development Programme

The evaluation did not identify positive impact on learning outcomes (S3.1). Statistical tests for teacher training leading to improved learning outcomes were not significant (S3.2). There is some suggestion that the ability of teachers to assess and monitor pupils learning is not adequate and so test results may not be truly reflective (S3.3). Mismatch of languages between home language and language of instruction may also be a barrier to learning (S3.9).

3.3 Non-Formal Learning Centres (NFLCs)

The evaluation looked at learning outcomes from running the programme with and without additional coaching for the trained facilitators. The programme with and without coaching both improved learning compared to a control (S4.1). The assessment showed that coaching did not improve learning outcomes (S4.2) or the quality of learning facilitators' instructional practice (S4.3).

3.4 Reading and Numeracy Activity (RANA)

Learning gains were identified from the beginning and end of the school year (S5.1). Each year has shown additional improvements in learning outcomes suggesting that familiarity with the project is improving results (S5.2). Improved learning outcomes suggest that pedagogical improvements may have been successful, such as adding review periods and training teachers in classroom management techniques (S5.3). Teachers were sensitised to the importance of using the full class period for lessons and this has been observed in practice (S5.4).

3.5 Northern Education Initiative Plus

The evaluation suggests factors limiting impacts on learning outcomes: high rates of absenteeism among teachers and students; insufficient time on task; and low capacity and motivation of teachers (S6.1). Also identified at baseline were learners' lack of preparedness for school and infrequent visits by school supervisors (S6.2). Learning materials and teacher training were designed to address these issues and the midline review points to modest impact. Teachers are spending more time teaching and learning outcomes are higher (small but statistically significant correlations are found) (S6.3).

The programme was hampered by budget constraints and some teacher training was cancelled (S6.3).

Reading gains are thought to have been hampered by problems with access to teaching guides (S6.6).

4. Factors associated with teacher performance or effectiveness

4.1 GEP III

The evaluation found limited impact on the knowledge of trained teachers (S1.4). Curriculum knowledge deteriorated (S1.6). There was some improvement in pedagogical knowledge (S1.5) but this did not result in improved teaching practice (S1.15). Some improvement in teachers' ability to interpret English was found (S1.7). A slight impact on Hausa knowledge in IQSs was found (S1.9). There has been a decrease in the teachers' competency index from baseline to midline overall (S1.10). There was a weak significant effect found for RANA teachers' use of pupil-centred activities but only in IQSs (S1.11). No positive effects were found for teacher absenteeism (S1.12).

Motivation has improved on a composite index (S1.13). Self-efficacy was the largest contributor to the improvement in this index. Teachers who have been trained are more likely to think they can do their job well. However, analysis suggests improved motivation does not reduce absenteeism or lead to better teaching practices (S1.14).

The evaluation suggests that use of reading resources could improve teachers' comprehension skills (S.15) and using a phonics approach could improve their English (S1.8). The limited effectiveness of the training is suggested to be related to: low initial subject knowledge and pedagogical knowledge; language barriers; and the challenging class environment (S1.17).

4.2 Teacher Development Programme

Teachers' subject knowledge was not found to have improved (S3.4). The use of teaching practices that are considered effective has increased significantly (S3.5). Almost half of the teachers evaluated under the programme reported gaining skills in lesson planning (S3.8).

Teachers were found to enjoy improving their thematic and methodological knowledge (S3.10). However, overall, impact on commitment and enthusiasm was not identified (S3.6). The programme believed that improved pedagogy would lead to greater levels of motivation (S3.7). In reality, there are other factors such as salaries, working conditions, lack of improvement in learning outcomes all affecting motivation. Teachers reported that pupil attendance and learning progress is the biggest driver of motivation (S3.11). Qualitative research found that teachers had deeper motivations also of 'being part of the children's paths' and 'producing leaders of tomorrow' (S3.12). Excessive workload was identified as negatively affecting motivation levels (S3.13).

4.3 Reading and Numeracy Activity (RANA)

Teacher value added (VA) measures teachers' performance by learning growth achieved for their pupils (S5.5). VA scores are found to be higher when a teachers' head or lead teacher attends cluster meetings (S5.6). Teacher age, experience, class size, or lesson duration are not indicative of teacher quality. Teachers' employment status and salary appear to correlate with

teacher performance (S5.7). Teachers' education is also a significant predictor of teacher performance (S5.8). Female teachers tend to outperform male teachers (S5.9).

4.4 Northern Education Initiative Plus

The application of teaching best practices remains low (S6.7). Weak systemic government are among contextual problems suggested to hamper results (S6.8). Other problems suggested: low foundational level of learners, inadequate training, lack of resources (S6.9).

The midline report recommends increased training, regular coaching and mentoring, and other teacher professional development support (S6.4). Teachers have successfully improved students abilities in decoding and encoding texts (S6.5).

5. Relative value and effectiveness of different types of teaching and learning materials

5.1 GEP III

There was no reference made to the use of audio visual materials.

Access and use of teaching and learning materials in general has increased but use is sporadic (S1.18). The evaluation found that most IQSs received the teaching and learning materials but instructions were not given or were not clear (S1.19). Materials were also not aligned with the whole-language approach taught in training (S1.20). Hausa based materials are not suitable for areas speaking local languages.

An expert review highlighted some positives finding the learning materials distributed by the programme as well designed, simple and using appropriate language (S1.20).

It is recommended the materials focus more on local sociolinguistic situations and make stories relevant to the daily lives of the pupils (S2.2). See S2.2 in the appendix for detailed discussion on language and appropriateness of learning materials.

A Hausa language series called *Aqua Green* are commended in the evaluation as well-designed (S2.3). They work as an integrated learning unit that match letters, figures, words and pictures. They provide a simple inventory of objects and events matching local environments. Instructions for use were thought to be inadequate (S2.4).

5.2 Teacher Development Programme

Teachers in the study had good access to video and audio teaching materials (S3.14). Teachers had a mobile phone loaded with content for self-study (the Trainer in the Pocket (S3.19)). Later SD cards were provided for teachers to use in their own phones and amplifiers were provided to broadcast to the class.

Teachers reported finding audio visual materials useful and enjoyable (S3.15). However, the researchers did not observe teachers using these materials in classes (S3.16). Limitations of broken batteries and problems charging devices need to be addressed (S3.17). Teachers also noted a shortage of training for the audio visual material (S3.20).

Textbooks were present in just over 50% of the classrooms (S3.18). Teachers noted this shortage as a problem (S3.20). When teaching and learning materials are kept in the head teacher's office, access can be a problem (S3.21). Delays in printing the programme material were a problem with some materials not being distributed until three years into the programme (S3.22).

5.3 Reading and Numeracy Activity (RANA)

The programme did not include audio visual materials. Learning materials in general were found to be lacking in classrooms evaluated (S5.11).

5.4 Northern Education Initiative Plus

Audio visual materials were not part of this project. Timely availability of teaching materials has been hampered by inadequate capacity to procure, store, and distribute materials to local areas, there are lack of resources for delivery, poor communications between officials organising distribution (S6.10).

6. Teacher development systems

The evaluations reviewed for this report had little discussion of the merits of different teacher development systems.

6.1 GEP III

Master trainers and School Support Officers conduct on-site monthly supervisory support and monitoring visits as part of the GEP III programme (S2.5). The success of peer-mentoring, supervision and Master Trainers is suggested to depend on a number of factors (S2.6):

- Teachers being literate in Hausa
- Weekly peer meetings being feasible given the context of the school (for example, teachers being able to meet regularly with other teachers given the size of the school or its location)
- Lead Teachers having the capacity, motivation and availability to lead the meetings
- Master Trainers and government staff having the resources, competencies and incentives to conduct effective supportive supervision

In one state, it was noted that Master Trainers had been trained but had not yet begun to train the teachers, so assessment of this had not been carried out (S2.7).

6.2 Teacher Development Programme

This programme used a cascade training model with a teacher development team receiving training to then convey to teacher facilitators (S3.23). The teaching facilitators reported mixed experiences of the training depending on the state (S3.24). There was feedback from the teaching facilitators and teacher development team that the programme had helped them to realise the importance of learning outside of just using a blackboard and the importance of praise (S3.25). The training was rolled-out differently in different states so it is difficult to get a clear picture (S3.26).

6.3 Reading and Numeracy Activity

Master Trainers were a part of this initiative but their role in training, coaching or mentoring was not described in detail in the evaluation. Master Trainers and School Support Officers were noted to have trained approximately 800 educators who then received monthly visits (S5.12). RANA Master Trainers and teachers received a 1-day training in Early Grade Reading Concepts (S5.13). The report describes the role of Master Trainers and School Support Officers in the collection of evaluation data (S5.14).

7. Changes in management at state or local government level

7.1 GEP III

The programme has no formal capacity development strategy though it is occurring through various trainings. Government stakeholders reported being grateful for opportunities the programme has brought to increase capacity (S1.21). Staff working in Education Management Information System departments and those working on the Annual School Census particularly spoke of large capacity benefits (S1.22).

Stakeholder interviews suggest a problem where different projects are operating within the same state, which confuses the system with inconsistencies in approach (S1.23). There tends to be a focus on working with individuals rather than governors as a group (S1.24). The report calls for more equality in this area.

7.2 Northern Education Initiative Plus

Assessors for the Initiative were trained from state education institutions as part of the system strengthening agenda (S6.14). The approach was noted to be simple and informative with encouraging buy-in (S6.16).

System strengthening was not discussed further in the report. It was noted poor results may be down to weak systemic governance and inconsistencies from state actors that the Initiative were largely unable to address (S6.15).

8. Evidence leading to programme adaptation

8.1 GEP III

The facilitator training initially used the approach of the TDP but this was found to be unsuitable for IQSs and was adapted to use the whole-language approach and promote development of lesson plans and learning materials (S1.27).

The evaluation design was adapted as changes were made to the intervention implementations (S1.26).

A key recommendation is the need to further adapt learning materials for the extremely low competency levels of facilitators (S1.25). In addition, standardised implementation is cautioned against due to diverse state contexts and customised adaptation is recommended (S1.28). Adaptation of training approaches is warned against however as ongoing revisions can create confusion (S1.29).

8.2 Teaching at the Right Level (TaRL)

The remedial learning programme reported on was built on a citizen-led household assessment, LEARNigeria (S8.1).

8.3 Reading and Numeracy Activity

Adaptations were made to the teaching approach, which had become 'less relevant' (S5.15), it was not clear where the evidence for this decision came from. The approach: 'I do' 'We do' 'You do'¹ was noted in the report to be more rigid and minimum teaching competencies were redefined so as not to be based on that approach.

9. Gender sensitive programming

9.1 GEP III

The evaluation concludes that the intervention is highly unlikely to have improved gender-sensitive class practice (S1.32). No difference in practice at midline compared to baseline is

¹ <http://www.evidencebasedteaching.org.au/the-i-do-we-do-you-do-model-explained/>

observed though difficult to measure (S1.35). The design and implementation of the programme lacks clear mechanisms for this to be achieved and improved. Gender biases among facilitators prevail (S1.33). Training for facilitators does not address gender in teaching practice and focus on training females is not prioritised (S1.34). The evaluation recognises an increased need for facilitators to acknowledge their role in improving girls' participation in class (S1.36). Facilitators seem to be aware of gender-sensitive class practices but this has not yet constituted a shift in attitudes and behaviour (S1.37).

Female teachers are found to be underrepresented, particularly in IQSs where around 5% of teachers are female (S1.31). Female enrolment has not improved (S2.9). There has been a slight increase in female representation within Centre-Based Management Committees (CBMCs) (S2.10). Only 48% of CBMCs with a written development plan report having girl-friendly practices (S2.11). Separate functioning toilets for girls have increased slightly (from baseline to midline) and are more likely where a female teacher is present (S2.8).

Information on gender-sensitive factors is noted in the GEP III evaluation as difficult to define and hard to reliably measure (S1.30). The report recommends clarification of concepts and ambitions with regards to making the school environment more girl friendly (S1.38).

Increased learning was not identified for girls or boys (see section 3).

9.2 Teacher Development Programme

Upon review of the endline survey, gender-sensitivity does not seem to be a focus for this programme. The gender learning gap present at the beginning of the programme has persisted in English, Mathematics and Science (S3.27). Stakeholders interviewed mentioned the success of a separate UNICEF programme running alongside that focusses on increasing girls' enrolment (S3.28). Access to sanitation for girls is noted as problematic (S3.29).

9.3 Non-formal Learning Centre

Gender-sensitivity was not a focus of this programme. Baseline differences between girls and boys academic attainment differences were not found and neither treatments in the programme affected boys and girls differently (S4.4). Programme organisers have teamed up with TEGA-Girls Effect to explore the gender dimension of the endline evaluation (S4.5).

9.4 Reading and Numeracy Initiative

In the RANA programme, from year 1 to year 3, there was a 34% increase in the number of girls who could read 31 words per minute (S5.16). This seems to be as a result of the general programme rather than being tied to a specific activity that enabled girls to learn.

9.5 ENGINE II Learning needs assessment

Streaming 'out of school' girls was identified as important for learning. Identifying whether girls needed to be in basic, post-basic or continuing education classes (S9.1).

9.6 Research on Female Teachers

Data was analysed from the Nigerian Education Management Information Systems (NEMIS) across eight states in Northern Nigeria and learning assessment data from RANA in Katsina and Zamfara states. The secondary data was supported with qualitative primary data from Katsina, Zamfara, and Niger. The research identifies a positive effect of the presence of a female teacher on enrolment of girls (S10.1). The amount of girls enrolled increases with more female teachers present (S10.2). Discussion around causality was not included. The presence of female head teachers is not as impactful in terms of enrolment as having female teachers in the classroom (S10.3). Female teachers are found to be more impactful in improving girls' learning outcomes than male teachers (S10.4). This result is also found for boys (S10.5).

10. Approaches to and instruments for measurement

10.1 GEP III

Teacher knowledge was evaluated in three areas: curriculum knowledge, subject knowledge, and pedagogic knowledge (S2.12). A knowledge and skills test comprised of 34 questions including multiple choice, short response and long response (S2.13). Marking and cross-checking was carried out by two senior education specialists (S2.14). Responses were used to develop subscales for eight areas of teachers' knowledge and skills (S2.15). High levels of missing data were noted as a concern and impeded the capacity to draw strong conclusions (S2.16).

An instrument to measure teacher motivation was developed based on existing instruments and a range of theoretical backgrounds (S2.17). An overall teacher motivation index was calculated as the average of the scales: self-efficacy; interest and enjoyment; pressure and tension; and effort and importance (S2.18). The sub-scales and composite index are noted to be difficult to interpret (S2.19).

Pupil literacy assessments were constructed using six steps. 1) Clarifying constructs: the GEP III uses the broader constructs of literacy rather than reading (S2.20). 2) Targeting: tests were designed drawing on existing EDOREN tests to measure skills below, at, and above the skills assumed to have been reached (S2.21). In low-income countries, testing is often criticised for being too high level. Rasch modelling is used to link difficulty levels between baseline and midline to cover a wide range of pupil knowledge (S2.22). 3) Administration: conducted with a tablet and a booklet when necessary (S2.23). 4) Psychometric analysis (S2.24): the psychometric metric properties of test items were tested, items that did not perform well were removed to avoid biasing results. A psychometrician used software to undertake this. 5) Proficiency bands: informed decisions were made regarding the standards of learning (S2.25). 6)

Secondary analysis: the primary analysis produces two variables for secondary analysis, a scale score and a proficiency band (S2.26).

10.2 Teacher Development Programme

A quantitative survey included interviews with head teachers, lesson observations, and a teacher development needs assessment (S3.30). Pupils were tested on English, mathematics and science. The tests were designed using the Nigerian primary syllabus (S3a.3). The results are analysed to account for differences in difficulty between questions (S3.31). The psychometric properties are analysed using the Rasch model (S3a.4).

Some of the measures rely on reporting by teachers and head teachers which may cause bias. Particularly when measuring motivation (S3a.1). Measures of teachers' behaviour rely on observations where teachers may adjust their behaviour because they know they are being observed (Hawthorne Effect) (S3a.2).

10.3 Non-Formal Learning Centres (NFLCs)

A background information questionnaire was used to collect student demographic characteristics (S4.6). The Early Grade Reading Assessment is used to assess literacy and has been validated in Nigeria (S4.7). Early Grade Mathematics Assessment was used to assess numeracy (S4.8). Children's socio-emotional skills were also assessed (S4.9).

The Teacher Classroom Observations tool was used and results tested well for reliability (S4.10).

10.4 Teaching at the Right Level (TaRL)

The TaRL document reports on results from the 2017 LEARNNigeria survey (S8.1). The approach to measuring literacy and number recognition was not discussed in the source for this report.

10.5 Reading and Numeracy Activity (RANA)

Master Trainers visit schools every month conducting interviews, learning assessments and classroom observations (S5.17). Master Trainers were themselves trained to carry out the monitoring (S5.18). Early Grade Reading Assessment (EGRA) and Early Grade Mathematics (EGMA) testing was used to assess children's learning. The report acknowledges that these as with all assessments are subject to some degree of error (S5.20). A value added model is used to distinguish between schools and teachers based on learning outcomes (S5.19). This can improve targeting and help to investigate the correlations of pupil learning. An error-in-variables regression model is used to account for measure reliability (S5.21). The report notes the success of the RANA monitoring and evaluation (M&E) system as being straightforward and relying on standard M&E tools (S5.22). The report authors emphasise the importance of adding pupil-level

longitudinal tracking within year as particularly informative and often missing from early grade assessments (S5.23).

10.6 Northern Education Initiative Plus (S6)

EGRA is measured at baseline and for subsequent cohorts to estimate the effects of the Northern Education Initiative Plus activity (S6.17). The Initiative conducted a four-day workshop to review the existing EGRA instruments, developing and piloting different versions (S6.18). The pilot conducted following a design in the EGRA Toolkit (S6.20). The Initiatives M&E team reviewed the findings to decide on the best version.

The Initiative note the limitation of carrying out training for the data collection in two different states and took measures to ensure as much harmonisation as possible (S6.19).

10.7 ENGINE II

A Learning Needs Assessment for children was carried out by trained facilitators and programme officers. The instruments are not detailed in the report reviewed for this helpdesk.

10.8 Research on female teachers

The research uses RANA data and NEMIS data to look at children's learning.

11. References

The references are provided in Section 2: The evaluations.

Suggested citation

Bolton. L. (2019). *Evaluation learning review for DFID Nigeria*. K4D Helpdesk Report. Brighton, UK: Institute of Development Studies.

About this report

This report is based on fifteen days work. The K4D research helpdesk provides rapid syntheses of a selection of recent relevant literature and international expert thinking in response to specific questions relating to international development. For any enquiries, contact helpdesk@k4d.info.

K4D services are provided by a consortium of leading organisations working in international development, led by the Institute of Development Studies (IDS), with Education Development Trust, Itad, University of Leeds Nuffield Centre for International Health and Development, Liverpool School of Tropical Medicine (LSTM), University of Birmingham International Development Department (IDD) and the University of Manchester Humanitarian and Conflict Response Institute (HCRI).

This report was prepared for the UK Government's Department for International Development (DFID) and its partners in support of pro-poor programmes. It is licensed for non-commercial purposes only. K4D 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 DFID, K4D or any other contributing organisation. © DFID - Crown copyright 2019.

12. Appendix

Girls Education Project Phase 3 (S1 and S2)

S1.1. p37- 38	RANA has had a small significant impact on pupils' average Hausa scale scores. This impact on the average score is not large enough to have resulted in a shift in the proportion of pupils falling into each performance band. While this small impact is observable for girls and boys equally in IQSs, in public schools there is no evidence of a significant impact of RANA on boys. The only significant impact is observed among girls there, but it is very small. RANA has had a small significant impact on the average English score in public schools, but not in IQSs. These effects that are attributable to RANA in public schools are more accentuated for girls than boys.
S1.2. p40	Those assessed as at the level of pre-literacy went from 97% at baseline to 90% at the Midline survey. Those at emerging literacy levels went from 1% to 6% and those at basic literacy levels went from 2% to 4%.
S1.3. p65	Overall, this correlation analysis finds very few factors that are significantly influencing changes in literacy outcomes, and those that are significant are generally small in magnitude. This is likely to be a result of the limited variability that we observe in both learning outcomes and key factors of interest. For instance, teacher competency has not improved on several indicators and any observed improvements have been very small. This can therefore help explain why we do not observe a significant correlation between changes in teaching knowledge and practice and changes in pupils' learning outcomes.
S1.4. p45	Overall, there appears to be a very limited impact of RANA on the knowledge of trained teachers. RANA appears to have improved teachers' skills in Hausa-based instructions in IQSs, but not in public schools.
S1.5. p47	Teachers scoring zero for pedagogical knowledge went from 90% to 65%. How much higher than zero was not detailed.
S1.6. p47	Teachers' subject knowledge has not improved significantly, curriculum knowledge levels have deteriorated significantly.
S1.7. p47	RANA has had a significant impact on teachers' skills in 'interpreting English words and phrases' and on teachers' English comprehension skills. RANA teachers are 6

	percentage points more likely to be competent in this skill compared to teachers in control schools.
S1.8. p48	Even if they are in a different language, access to and use of materials like storybooks and other reading resources would have improved teachers' comprehension skills. The focus on teaching by using a phonics approach could also have disproportionately improved their English.
S1.9. p48	While there is no overall impact on Hausa, we find some evidence that RANA has had an impact on Hausa knowledge in IQSs. Teachers in RANA IQSs are about 19 percentage points less likely to show no evidence of this skill (bottom band) and 12 percentage points more likely to be competent in this skill (top band) than teachers in control IQSs. No such impact is observed in public schools. There is no impact of RANA on Hausa comprehension skills in either public schools or IQSs.
S1.10. p51	Overall, treatment teachers' performance on the teacher practice composite index has decreased slightly but significantly between baseline and midline.
S1.11. p52	There is some evidence that RANA has had a small impact on teachers' instructional skills in active early grade learning, based on the teacher practice composite index, although this effect is only weakly significant. When looking at the activities that make up the composite index, there is robust evidence that RANA has had an impact on the teachers' use of pupil-centred activities in the classroom. The effect of RANA on the overall teacher practice composite index is larger and statistically significant only in IQSs.
S1.12. p56	Teacher absenteeism seems to have declined in IQSs but not in public schools. It is important to point out that this measure is based on self-reported attendance
S1.13. p57	Motivation has increased between baseline and midline among teachers in RANA schools. Specifically, only 30% of the teachers at baseline had the level of motivation that the average teacher has at midline. Examining the different sub-scales that make up the motivation index used in the analysis, there have been increases in interest and enjoyment, pressure and tension, and effort and importance. However, the increase in the overall motivation comes primarily from a large increase in the self-efficacy scale. Interestingly, self-efficacy may be the aspect of motivation that is most amenable to training, because teachers who have been trained are more likely to think they can do their job well. This could therefore be the result of midline teachers in our treatment group having attended RANA training.
S1.14. p58	analysis also suggests that improved motivation does not substantially reduce absenteeism, nor does it lead to teachers using better teaching practices at midline

S1.15. p77	The subject and pedagogical knowledge among facilitators teaching in GEP3-supported IQSs has improved slightly. Overall, this has not translated into more effective teaching practice. GEP3's contribution to more effective teaching in supported IQSs is overall likely to be limited, although we observe some improvement in pedagogical practices and use of teaching and learning materials that were promoted during GEP3 training. In general, GEP3's IQS support has only made a minor difference in expanding the number of competent facilitators and improving effective teaching practice.
S1.16. p78	GEP3's contribution to more effective teaching in supported IQSs is likely to be limited, although GEP3-promoted teaching practices and teaching and learning materials are observed in some schools. GEP3's main mechanisms for improving effective teaching are by enhancing facilitators' knowledge and skills and by providing and promoting the use of more teaching and learning materials.
S1.17. p81	The overall limited influence of the training is likely to be related to the very low subject and pedagogical knowledge levels of the facilitators, language barriers, the challenging class environment, and the fact that the revised training had not been implemented in all LGAs
S1.18. pxi	While the use of teaching and learning materials during classroom instruction has increased, their use remains sporadic. The percentage of classrooms with access to and use of any teaching and learning has significantly increased since baseline. However, specific materials such text books, Hausa materials or locally made materials are still only observed in a small minority of classes. Also, the teaching and learning materials distributed by GEP3 in 2016 were rarely used and are frequently not available anymore. Effective use of these materials has likely been affected by lack of instruction and guidance on how to use the materials and the absence of their integration in the IQS facilitator training.
S1.19. p75	Teaching and learning materials reached most IQSs, but with delays and without clear instructions. According to the survey data, 98% and 84% of IQSs in Bauchi and Niger, respectively, received a box of Hausa-based teaching and learning materials. While UNICEF confirmed the distribution of the materials to state government agencies between April and June 2016, there were delays in the distribution to the schools. Instructions regarding how the materials should be used were either not given or were not clear.
S1.20. p91	An expert review of the GEP3-distributed teaching and learning materials package assessed the materials as well designed, simple, using appropriate Hausa language, aimed at early grade pupils, and as a promising teaching resource in the hands of

	<p>competent teachers.² However, the review also observed that the lack of instructions could make them challenging to adapt for use in a school context such as that of the IQSs. Furthermore, the distributed teaching and learning materials were not aligned with the whole-language approach promoted during the adapted training, which, without further instruction, is likely to affect their use. Also, the Hausa-based materials are not appropriate for non-Hausa-speaking environments, like Nupe communities in Niger, and can even have negative consequences for literacy learning, particularly without their use being accompanied by sufficient instruction (see Error! Reference source not found.). The promotion of the use of materials from the local environment, on the other hand, is more appropriate for non-Hausa communities, and is perceived by facilitators as useful in regard to improving pupil engagement.</p>
S1.21 p17	<p>A striking feature of GEP3 is that there is no formal capacity development strategy. Clearly, capacity development is taking place through various forms of training, as well as on-the-job coaching, and most government stakeholders are grateful for the opportunities they have had.</p>
S1.22. p17	<p>In particular, staff in Education Management Information System (EMIS) departments and working on the ASC state that their capacity has increased hugely, to the point that they can deliver a high-quality product which is reliable and used by decision makers. This capacity is confirmed by UNICEF and NEI+, both of which have run validation exercises in parallel to the ASC and found the margin of error to be very small. The decision to stop using external enumerators and to train head teachers to gather data has increased the reliability, as well as developing an additional skillset.</p>
S1.23. p17	<p>In Bauchi, where there are other projects, one stakeholder observed that all projects aimed to support the system but they questioned whether they were, in fact, confusing the system in the absence of a government coordination mechanism capable of managing duplication or inconsistency in approaches. Another commented that donors are much more interested in impact than the government and said: <i>'government is like a hungry lion – it will take anything it is offered by donors'</i>. Although there are some attempts by donor-supported projects to coordinate among themselves, a strategy to approach the Governor as a group rather than as individuals has not yet materialised.</p>
S1.24. p17	<p>Although there are some attempts by donor-supported projects to coordinate among themselves, a strategy to approach the Governor as a group rather than as individuals has not yet materialised.</p> <p>Capacity can also be depleted where there are competing projects as all projects are seeking to work with those individuals in government who demonstrate capacity and commitment. This is a point that has been raised in Sokoto by another project; this was confirmed by a government focal person, who observed that the SMOE relied</p>

² The report of the teaching and learning materials expert review are included in Chapter 7 of the Technical Report.

	heavily on those staff whose capacity has been built through GEP3. This has led to jealousy among other officials who have not had the same opportunities for capacity development and feel that they are further penalised by being passed over for general opportunities.
S1.25. pxii	The facilitator capacity building process and any distribution of teaching and learning materials need to be further adapted to the extremely low competency levels of the facilitators.
S1.26. p6	The evaluation was conducted broadly in line with its original evaluation design, as presented in the GEP3 Evaluation Framework, although some methodological adjustments were made to adapt the evaluation design to changes in intervention implementation, lessons learnt from baseline data collection, and changing information needs of GEP3 stakeholders.
S1.27. p73	The approach and content of the IQS facilitator training was revised during the evaluation period. From August 2015 until September 2016, facilitators were trained according to an approach used by the TDP. Because the TDP's content and approach was not considered sufficiently fit for an IQS context, the facilitator training was adapted in the last quarter of 2016. The adapted training changed the training cycle, providing a staggered training of four termly multi-day workshops. Furthermore, the adapted training introduced another approach to Hausa teaching—the whole-language approach—and promoted the development of lesson plans and teaching and learning materials.
S1.28. p117	Although there is general acknowledgement that diverse state contexts require different approaches to implementation, in practice, this is difficult to achieve. The structure of the state teams, with recruitment aligned according to the four outputs of GEP3, tends to result in somewhat standardised implementation. In order to maintain and increase relevance, GEP3 implementation needs to be able to demonstrate a greater degree of customised adaptation.
S1.29. p120	While adaptation of the approach will be necessary given the diverse and complex context of IQSs, ongoing revisions of the teaching approach promoted at the training possibly creates confusion among facilitators that are already struggling to understand the training content. A stable intervention design and implementation approach is needed before scale-up.
S1.30. p64	As also discussed at baseline, it is challenging to achieve a reliable measure of the girl-friendliness of schools, as information on gender-sensitive factors is difficult to define and collect. The information that was obtained on this for our early learning sample cannot therefore be considered exhaustive.

S1.31. p64	Female teachers continue to be underrepresented in the sample schools, with the proportion of female teachers still below 10%. There is, however, a big difference between public and IQSs, with the former having almost 50% of female teachers and the latter only around 5%.
S1.32. p77	It is very unlikely that gender-sensitive class practice has improved and that GEP3's intervention has made a difference. GEP3's design and implementation lack clear mechanisms for how gender-sensitive class practice is to be achieved and improved.
S1.33. p78	On the surface, gendered differences in facilitators' pupil engagement appear limited. However, the qualitative classroom observations show gendered differences in the quality of this engagement. Therefore, girls are still marginalised in the quality of classroom interaction and engagements. The qualitative research indicates that facilitators do not demonstrate any new gender-sensitive practices in class beyond those observed at baseline. While facilitators generally verbally express positive attitudes toward girls' education, gender biases among facilitators prevail when probed deeper.
S1.34. p86	GEP3's ToC and strategy papers promote the use of gender-sensitive teaching, although neither the concept of gender-sensitive teaching nor the mechanisms for how to achieve it are clearly defined. ³ The training sessions of the first two rounds of the adapted facilitator training did not address gender in teaching practice; neither has the GEP3 training been able to prioritise training of female facilitators. Therefore, both in design as well as implementation it remains unclear how teaching is meant to become more gender-sensitive.
S1.35. p86	differences in pupil engagement appear to be limited, no differences in practice compared to baseline are observed and the quality of interaction and engagement still shows gender bias. Based on the quantitative data, overall no strong gendered differences are observed in the way teachers engage with girls or boys within the classroom. ⁴ However, gender biases in teaching practice are difficult to measure through quantitative methods. As observed at baseline and again at midline, facilitators perform gender-sensitive practices in a tokenistic way—in line with development projects' sensitisation efforts and influenced by social desirability during observation (knowing what one should do). ⁵ The qualitative research indicates that facilitators do not demonstrate any new gender-sensitive practices in class beyond

³ See the GEP Midline Evaluation Technical Report, Chapter 4, for a review of gender-sensitive teaching references in the GEP3 strategy documents.

⁴ Gendered pupil engagement is observed in terms of asking open or closed questions differently for girls and boys, asking girls and boys to assist in the lessons differently, using a girl child's name versus a boy child's name, or whether the teacher interacts with groups of boy pupils in the class more than groups of girl pupils.

⁵ For example, during one qualitative classroom observation the facilitator asked only boys during the first 15 minutes, then said, 'Oh, I should ask girls now, otherwise they will think I only ask boys!'

	<p>those observed at baseline (mainly around participation: asking every other girl/boy questions). While the ‘quantity’ of the facilitators’ engagement with girls versus boys may be similar, the ‘quality’ of engagement is different. For example, facilitators ask boys more open-ended questions, and girls more close-ended or follow-up questions. In cases where girls are asked open-ended questions these are often questions that have already been asked of boys or the facilitator does not engage fully with the responses as he does with boys. Similarly, when it comes to reading stories from the board, boys are asked first in the majority of cases, and girls second (once something has already been demonstrated by a boy).</p>
<p>S1.36. p86</p>	<p>While facilitators generally express positive attitudes toward girls’ education, gender biases and prejudice among facilitators prevail. Similar to baseline, facilitators interviewed as part of the qualitative research all state that it is important for girls to go to school, and state that this is as important as for boys.⁶ However, when probed further as to why girls should study facilitators tend to stress gendered expectations about girls’ future roles in the household and society: for example, the importance of women being able to help their children with homework, or being nurses (so that women can be treated by women). While some facilitators in Bauchi indicate that they want to focus more on girls than on boys because they are aware that girls will leave school earlier (drop out), others stress that girls are ‘shy’ and therefore do not participate in class, not acknowledging their own role in including girls, nor giving reasons for why girls may be shy, but placing responsibility for this on the individual girl child.</p>
<p>S1.37. p86</p>	<p>The prevalence of gender prejudice among facilitators is further evidenced through the findings of the facilitator survey, in which facilitators were asked ‘<i>why some people think that a woman should acquire her knowledge from a man and a man must have more knowledge than a woman</i>’; followed by whether the respondent agreed with the statement. Almost all surveyed facilitators had an understanding of the varied reasons driving gender prejudices that can limit and curtail educational equality for girls and women – mentioning, first of all, reasons regarding an intrinsic greater intelligence of men over women, or referring to men’s greater abilities compared to women. This was followed by reasons related to constraints women and girls face societally as a result of gender norms, such as marriage, lack of women’s mobility, and lower educational chances leading to their subordinate educational learning to men. Not only do facilitators have an understanding of the reasons driving gender prejudice, 65% agree with the prejudice. Consequently, though facilitators are aware of gender-sensitive class practices, as discussed above, and acknowledge the importance of girls’ education, this does not constitute a deeper shift in gender attitudes.</p>

⁶ There is a strong indication of social desirability bias in these data, with facilitators becoming quite uncomfortable with regard to these questions in the majority of cases.

S1.38. p119	Recommendations: clarify concepts and ambitions about gender-sensitive teaching and making the school environment more girl-friendly.
----------------	---

S2.1. p198-9	In particular, teacher competency has not improved between baseline and midline on several indicators, and where it has, the improvements have been very small. This can help explain why we do not observe a significant correlation between changes in teaching knowledge and practice and changes in pupils' learning outcomes.
S2.2. p291	Hausa pupils, who are barely able to recognise the Hausa alphabet before their first lesson in public schools, are likely to struggle with the RLP materials as a package. Combining phonetic and phonemic awareness in a phonics environment, with its universal applicative assumption, even in tone language environments, may constitute a source of frustration for the pupil. Of course, there is value in the teacher reading to the pupils, but even here, the materials, especially P2 and P3 stories, will require further editorial processes before the teacher is able to read the stories with fluency, speed, and contextual effect. The materials should also be more focused on the states' sociolinguistic situation, the pupils' emergent literacy, as well as the connections between the stories in the books and the daily lives of the GEP3 pupils. Equally important is the need to further simplify and shorten the stories. Yet, even after overcoming these challenges, only an intensive and continuous training of GEP3 teachers can overcome the high level of expectation of pedagogic competence required by the RANA teacher's guide. A major shift in the Hausa orthographic system will also be required to deal with the issue of tone and vowel length, and how these may be integrated in future literacy and numeracy activities. RANA will need to discuss these issues with the relevant state partners to achieve a more holistic solution to a problem that could have a significant effect on Hausa language testing.
S2.3. p295	<i>Aqua Green</i> materials have positive elements to recommend them, particularly their excellent design, an integrated learning unit that matches letters, figures, words and pictures, as well as its simple inventory of objects and events closely matching local resources outside the school environment. The next step is to expand the materials, adapt them further for language use and choice, as well as to support IQS facilitators (allowing for their diverse qualifications and backgrounds) with a methodology that addresses how to use the materials in the classroom. This is especially important if the materials are to fit into the whole-language literacy approach espoused by IQSs.
S2.4. p294	The <i>Aqua Green</i> materials follow the Hausa alphabetic inventory sequence from <i>Aku</i> 'parrot', to <i>Mage</i> 'cat', to <i>Zomo</i> 'rabbit'. At the end, exercises requiring the learner to supply a missing letter are provided. Similarly, both the numerals and writing books are sequentially ordered, with numbers running from <i>daya</i> '1' to '100', and letters to be traced starting from A and running to Z. In the numerals book, each number is accompanied by equivalent figures (= 1-100) of pictured item(s), e.g. '43', <i>Arba'in da</i>

	<p><i>uku</i>, '43 corns' <i>masara guda araba'in da uku</i>. In this way, pupils' prior mother-tongue vocabulary is harnessed throughout the book, hence the usefulness of the books as an initial resource at Stage 1 of IQSs. Some form of learner autonomy is built into the books because of the integrative use of letters, words, figures, and pictures as learning units. Successful implementation of these materials will depend on the provision of differentiated and intensive training to facilitators in the language of the immediate community (not necessarily Hausa), as well as a flexible methodology that is sufficiently long and contextualised to harness the pupil's prior listening and speaking skills in their home language.</p>
S2.5. p68	<p>Master Trainers conduct on-site monthly supervisory support and monitoring visits. SUBEB and/or Local Government Education Area (LGEA) staff, called School Support Officers (SSOs), also conduct School Support Visits (SSVs).</p> <p>During these SSVs, Master Trainers and SSOs observe how well the peer-mentoring is working; provide mentoring to trained teachers, lead teachers and head teachers; conduct classroom observations to assess the fidelity of implementation to the scripted lesson plans; and collect other real-time monitoring information (RANA, 2016a, 2016c).</p> <p>RANA state officers also conduct SSVs to a random subset of schools each month, while RANA country office staff conduct some SSVs each quarter (RANA, 2016c).</p> <p>Master Trainers also conduct monthly cluster-based meetings with head teachers and Lead Teachers.</p>
S2.6. p71	<p>To optimise teacher learning, the project envisioned a teacher development approach based on progressive professional development and school-based peer-mentoring, complemented by supportive supervision and monitoring by Master Trainers and government staff. The success of this approach depends on teachers being literate in Hausa, weekly peer meetings being feasible given the context of the school (for example, teachers being able to meet regularly with other teachers given the size of the school or its location), Lead Teachers having the capacity, motivation and availability to lead the meetings, and Master Trainers and government staff having the resources, competencies and incentives to conduct effective supportive supervision.</p>
S2.7. p84	<p>By July 2017, Zamfara State had begun training Master Trainers to deliver the scale-up of the RLP, but had not yet rolled out the training to in-service teachers</p>
S2.8. p161	<p>a larger proportion of schools have separate functioning toilets for girls at midline, compared to baseline, particularly in public schools. Interestingly, schools that have at least one female teacher at midline appear to be somewhat more likely to have separate toilets for girls than schools with no female teacher: in public schools, 80% of schools with a female teacher have separate toilets for girls, compared to 69% of schools with no female teacher.</p>

S2.9. p161	At midline, there are 0.78 girls enrolled for every boy in P1–P3. This does not present a significant change compared to baseline. Schools in Katsina continue to have a higher ratio of girls to boys in P1–P3 (around 0.9), compared to schools in Zamfara (around 0.7), while public schools and IQSs have similar ratios (around 0.8). Trends in PTRs across all primary grades are comparable to those for P1–P3.
S2.10. p247	CBMC guidelines indicate that CBMCs should have a minimum of 12 members: nine of 60 CBMCs at midline have fewer than 12 members, compared to 12 of 52 CBMCs at baseline. Five CBMCs had fewer than 12 members both at baseline and midline; however, in four of these five cases, the total number of members has increased since baseline. Overall, these trends indicate that CBMC membership has improved between baseline and midline. The presence of female members, adult and child, has shown slight improvements between baseline and midline
S2.11. p258	Only 46% of Centre-Based Management Committees (CBMCs) with a written Whole-Centre Development Plan (WCDP) reported having any girl-friendly practices in their WCDP.
S2.12. p45	Adapted from Leach (2002) and Cogill's (2008) models of pedagogy, the knowledge model set out in the figure below conceptualises the knowledge that contributes to observable classroom practices. Within the model, there are three types of knowledge that a teacher draws on within classroom practice. Curriculum knowledge refers to knowledge of what should be taught to a group of students, knowledge of the national syllabus, understanding of the school and grade-level planning documents, and knowledge of the content of examinations. Pedagogic knowledge refers to knowledge of the learners in the setting, knowledge of how to provide the conditions that enable pupils to understand, and the selection of learning and assessment materials. Subject knowledge refers to knowledge of the essential questions of the subject, the networks of concepts, the theoretical framework, and methods of inquiry.
S2.13. p45	The teachers' knowledge and skills test was divided into three sections, collectively comprising 34 items, including multiple choice, short response, and long response items.
S2.14. p46	Teachers' responses to the items were marked and cross-checked using a marking matrix and scoring guide. The marking and cross-checks were carried out by two senior education specialists. One specialist is a previous director in Nigeria's education system and the other is an international education specialist
S2.15. p46	Teacher responses were used to develop eight areas of teachers' knowledge and skills: <ol style="list-style-type: none"> 1. syllabus knowledge; 2. ability to identify low performers;

	<ol style="list-style-type: none"> 3. ability to provide evidence for judgements and diagnose; 4. ability to build on student knowledge; 5. teachers' writing skills; 6. teachers' Grade 2 Hausa knowledge; 7. teachers' comprehension skills; 8. interpreting words and phrases; and 9. Hausa comprehension and writing skills.
<p>S2.16. p47</p>	<p>High levels of missing data were observed. The amount of missing data is a major concern. First, because it may indicate that many of the teachers who participated in the programme were not able to do many of the tasks asked of them. These tasks were intended to represent the core elements of what teachers should know and be able to do in order to teacher effectively, so the concern, here, is that the teachers may not be very skilled. This may suggest that the test was 'too difficult'. However, the test was designed around notions of what a minimally competent teacher should be able to know and do. It is probably, therefore, unhelpful to be overly concerned about the test's difficulty: the real issue is the competency level of the teachers. The second concern is that the range of analyses was limited due to the missing data, and hence the capacity to draw empirically strong conclusions was impeded.</p>
<p>S2.17. p48</p>	<p>Figure 1: Model of motivation for education evaluations</p> <pre> graph LR TS[Training and support] --> PK[better pedagogic / content knowledge / skills] TS --> IT[better interaction between teachers (5)] IT --> PK PK --> EL[external factors less limiting ('can do' / self-efficacy) (1)] EL --> ME[more motivated, engaged ('will do') (2 / 3 / 4)] ME --> BLO[better learning outcomes for students] BLO --> ME </pre> <p>Reproduced with kind permission of: EDOREN, 2018, p. 49</p> <p>Using this basic framework and drawing on a range of existing instruments with different theoretical backgrounds, an instrument was developed for measuring teacher motivation with the northern Nigerian context in mind.</p>
<p>S2.18. p49</p>	<p>An overall teacher motivation composite index was calculated as the average of the first four sub-scales (in table 8).</p>

S2.19. p49	The sub-scales and composite index can be difficult to interpret. A higher number in each index is associated with a higher level of the underlying construct, but we do not have benchmarks which would enable us to say a particular teacher is ‘very highly motivated’ or ‘insufficiently motivated’ in absolute terms. However, the scales allow us to make comparisons between groups and over time, so that we can say, for example, whether teachers in 2017 were more motivated than in 2015, or if teachers in rural areas are more motivated than those in urban areas. In order to make interpretation easier, we standardise the variables with reference to the <i>treatment group of panel schools at baseline</i> (using survey weights) in the early learning evaluation, and the <i>group of panel schools at baseline</i> in the IQSs evaluation.
S2.20. p43	At midline, the constructs to be measured through the learning achievement tests in the GEP3 evaluation are English literacy/language and Hausa literacy/language. While numeracy was measured at baseline, these skills are not covered in this round of the evaluation owing to the programme’s focus on early-grade <i>reading</i> and literacy. The GEP3 evaluation uses the broader constructs of literacy rather than reading, as discussions with stakeholders have highlighted that the programme expects children to improve learning in areas that fit into these broader areas of literacy—such as to understand texts, use reading to understand the world, draw inferences, and communicate in writing.
S2.21. p43	The assessments developed for this programme, at baseline as well as midline, do not assume that children will be able to read and write. Test items were designed or selected from existing EDOREN tests to measure skill levels below, at, and above the skills assumed to have been reached given the grade the child is attending. A major weakness in data measuring literacy in low- and middle-income countries is that assessments measure skills at levels that are too high for most of the children taking the tests.
S2.22. p44	As the assessments used for the evaluation are administered to a panel of pupils, the assessments need to be able to cover a wide variation in pupil knowledge (from the lowest performing pupils at baseline, to the highest performing pupils at midline – two years later). This places a great requirement on the test to credibly measure knowledge over a large range of difficulty, without making the test too long for young pupils. This constraint has been alleviated through the use of Rasch (item response theory) modelling. The tests use link items to link the difficulty of the baseline test to the midline test and to place the results onto the same scale (metric). The EDOREN metrics team has also placed this learning onto the same metric as has been used for ESSPIN English literacy, to allow for further contextualisation of the results observed in GEP schools.
S2.23. p44	Learning achievement tests used as part of the GEP3 evaluation use a set of assessment items or questions to assess children’s literacy and/or language knowledge and skills, based on a one-on-one individual oral interaction between a

	child and an enumerator. The test was administered using a tablet, while pupils were given a pupil booklet to mark their answers, or write, where necessary.
S2.24. p44	The use of Rasch modelling increases the amount of analysis required as more sophisticated techniques are used, rather than adding up a total number of items answered correctly in the test and converting the number into a percentage score. In this regard the evaluation team followed the following steps: the first step was to test the psychometric properties of the items to ensure they were useful measures of what pupils know. The second step was to remove any items that did not perform well and would bias the results if they were counted in the analysis. In a third step the team ranked the items according to difficulty. This was done by a psychometrician, using fit-for-purpose software. The software then also ranked pupils according to their ability and placed the pupils and the items onto the same metric. This is a probability model as pupils are placed on the scale according to the probability of a pupil answering the corresponding item correctly.
S2.25. p44	Informed decisions were then made regarding the standards of learning. For example, what skills do we expect pupils to have to consider them 'literate'? This defines the learning standard across the metric and creates a proficiency band. The percentage of children that fall within this range is then calculated. These proficiency bands were drawn through a benchmarking workshop undertaken with TDP, GEP3, Developing Effective Private Education Nigeria (DEEPEN), and ESSPIN stakeholders in November 2015 for the English literacy and numeracy assessments. For the Hausa assessment, the cut-off point was drawn to reflect the English literacy proficiency ranges as much as possible. In other words, the skills that fall within each proficiency range for Hausa are as similar as possible to the skills within the same proficiency range for English literacy.
S2.26. p45	The primary analysis produces two variables for each learning construct, per pupil. The first is a scale score. The scale score is a precise measure of where, along the achievement scale, the pupil sits. The scale score is useful for fine-grained analysis like multi-level modelling, regressions, and correlations. The second variable is which proficiency band the pupil falls within. This variable is useful for describing what learning pupils have achieved and how what children actually know and can do has changed, over the course of the evaluation.

Teacher Development Programme (S3)

S3.1. p73	the evaluation was not able to conclude that the programme had a positive impact on learning outcomes. First, the change in the treatment-control difference was not statistically significant.
--------------	---

S3.2. p75	There is no significant evidence for TDP training having an impact on pupil learning outcomes, irrespective of the subject. The estimation models implemented do not capture robust significant change in the test scores that could be attributed to TDP.
S3.3. p83	Qualitative analysis of interviews identified a perceived improvement in learning outcomes when asking pupils, teachers, head teachers and committee members different to analysis of test scores. This could be that respondents wanted to give socially desirable responses or may reflect teacher's weak ability to monitor and assess pupil's learning.
S3.4. p53	Teachers' subject knowledge has not changed in mathematics or science, and appears to have worsened since 2014 in English. There is no difference between schools that have received TDP intervention and those that have not.
S3.5. p56	The use of teaching practices considered effective has increased significantly between baseline and endline, and increased significantly faster in TDP than in control schools. This represents a significant positive impact of TDP on teaching practices and is robust to controlling for confounding factors.
S3.6. p68	Commitment and enthusiasm no difference between treatment and control.
S3.7. p68	The assumption underlying the programme's theory of change, that teachers' improved pedagogy, resulting from TDP training and support, would lead to greater levels of motivation, is not supported. There may be other factors, such as salaries, difficult working conditions, and a continued inability to bring about noticeably better learning outcomes, which still prevent teachers from having a sense of their own effectiveness, which might in turn make them more motivated. The absence of an effect of the programme on motivation may also help to explain the lack of impact on teacher absence. In interviews, many teachers did report that training improved their confidence in the classroom, but this appears not to have resulted in changes in measured motivation.
S3.8. p60	Close to half of treatment teachers (46%) reported having gained skills in lesson planning from the TDP training.
S3.9. p61	At endline 54% of observed treatment teachers used both English and Hausa when teaching Primary 1–3 lessons (for more see Chapter Error! Reference source not found.). That such a large proportion of teachers used both languages when teaching the lower grades is surprising given that nearly all pupils' home language is Hausa; the national policy on the language of instruction for Primary 1–3 is to use Hausa; and teachers' low English language skills. The use of English is likely to reflect the fact that most textbooks are in English, and a belief among teachers that English should be used. This practice may not be beneficial for pupil learning, as a

	mismatch between pupils' language at home and the language of instruction is a major barrier to early learning.
S3.10. p68	Teachers enjoyed improving their thematic and methodological knowledge: 'it is just like when you are in a room with a dim lamp and then the electric bulb is brought, of course it will be brighter', says one teacher, referring to the newly introduced TDP methods. Some teachers ranked the learning and researching aspects of their job as the most critical motivational driver. Teachers appear to be intrinsically motivated to attempt to use the methods and materials provided by TDP.
S3.11. p69	Many teachers perceived pupils' attendance to be among the most critical drivers of their motivation. Teachers in all states report that they are motivated by the learning, performance, and progress of pupils.
S3.12. p70	The qualitative research also finds evidence for a wider intrinsic motivation. A number of teachers reported being motivated by 'being part of the children's paths', 'producing leaders of tomorrow', and 'installing knowledge and values in children' (teachers in Jigawa, Zamfara, and Katsina). Teachers perceive their impact on pupils' lives to be substantial, as they frequently state that 'children taught by a teacher will remember the teacher', that the families of pupils value the teacher's contribution to raising their children, and that primary education is the foundation of all other education of a pupil in life (teachers in all three states).
S3.13. p70	Teachers' workloads can affect their motivation levels, in particular if they feel overwhelmed. Approximately one-third of the sample teachers consider their workload excessive. The main reasons given by these teachers are a shortage of teachers at the school, having too many classes to teach, too many pupils in their classes, and too many administrative and clerical duties.
S3.14. p63	The vast majority of the treatment teachers have access to the TDP videos on teaching methods (93%) and audio materials for use in the classroom (93%). Among the teachers who did not have access to the videos and audio, the main reasons given were that they had not been given to them and that they could not play them on their mobile phones.
S3.15. p63	In the qualitative research, most teachers and head teachers noted that they found the audio-visual teaching aids, such as the audio clips and videos, phone and amplifier provided by TDP to be useful inside the classroom, and recounted using them during lessons in a variety of ways. For example, they used the phone to show pupils images of nouns with which they were unfamiliar, such as 'lion' or 'tiger', and also to play videos to pupils by connecting the handset to the amplifier. Furthermore, teachers enjoyed using the audio-visual materials as these helped to keep pupils' attention and are useful for ensuring that pupils hear the correct pronunciation of English words.

S3.16. p63	However, the qualitative research team did not observe any teachers use TDP audio-visual materials during the lesson observations, and during the 247 lessons observed as part of the quantitative survey only three treatment teachers used audio. Taken together, this strongly implies that the audio materials produced by TDP for use in the classroom are currently not being used.
S3.17. p63	One possible reason is that across the states, teachers face difficulties using the amplifier due to broken batteries and problems charging the device. This is supported by the quantitative survey finding that although most of the treatment schools (93%) have a TDP amplifier, only 48% of these are working. The main reasons for the amplifiers not working are that they cannot be charged (71%) and that they are faulty (25%). To enable teachers to use the audio and video materials would require a solution that allows amplifiers to be charged and ensures that the financial means and technical support to repair faulty ones are available.
S3.18. p40	Across the three states, teachers and head teachers reported a lack of teaching materials. Teachers had a textbook in only 56–57% of the lessons observed in the quantitative survey, fewer than at baseline. A school in Zamfara reports not having received the relevant textbooks for their pupils, which leaves them having to borrow them from neighbouring schools. A head teacher in Jigawa notes: '[...] we don't have reference textbooks like Macmillan that we can use as reference to teach effectively so you cannot get examples to demonstrate.' By contrast, in over 90% of observed lessons in the quantitative survey, there was a functioning blackboard or whiteboard and chalk or a marker.
S3.19. p16	Cohort 1A teachers were each provided with a mobile phone loaded with audio-visual content – the Trainer in the Pocket – for self-study. For later cohorts this was revised to providing an SD card for use in teachers' own mobile phones. Two amplifiers were provided to each TDP school to allow teachers to play audio material in the classrooms.
S3.20. p20	Only 13% reported that the training covered use of audio-visual material during lessons, although this would be important for the use of TDP's audio clips in the classroom, for example.
S3.21. p22	materials are generally stored outside the classroom, generally in the head teacher's office, which may pose challenges for accessing and using materials
S3.22. P22	There were delays in developing and printing some materials in time for them to be distributed at the relevant TDP cluster training sessions. Delays in material distribution also affected the head teacher manuals for the school management and leadership training. The timeframe for developing the TDP materials at the start of the programme was expected to be six months, but later this was revised to two years. However, materials were still being developed and distributed three years after TDP

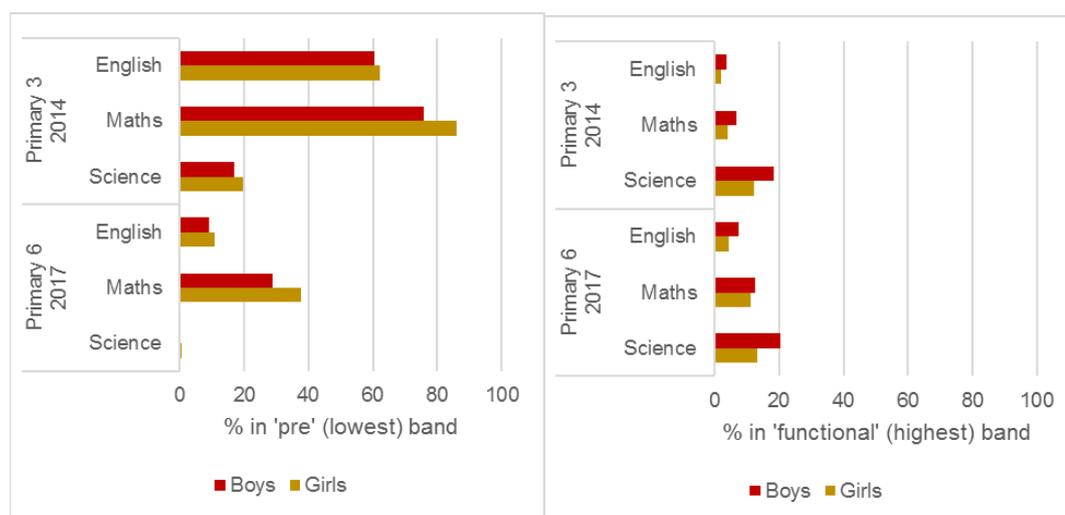
	<p>began. To mitigate these delays, photocopied versions of some materials were used as a temporary solution at cluster training sessions</p>
S3.23. p19	<p>Under the TDP cascade training model, TDTs were trained over two years, and received five days of training from international and national experts in Abuja, which they would then convey to the TFs during a three-day training session at state level.</p>
S3.24. p19	<p>Across the states, TFs felt differently about the degree to which they were prepared by their training to conduct cluster training sessions (Doyle <i>et al.</i>, 2016).⁷ In Katsina, there was a general sentiment that the quality of the training had declined, as had the quality of cluster training sessions in terms of arrangement and materials.</p> <p>'Sometimes they will just give you directive, go out and conduct so-so training, you are going to conduct so-so training without any preparations, no materials, nothing.' (TF, Katsina). On the other hand, TFs in Jigawa felt well-equipped by the training they received:</p> <p style="text-align: center;"><i>I also want to say, we are highly encouraged [...] because of the rigorous training which have been given from the TDTs. We are really stimulated, highly interested with regard to the training which has been given. The training which has been given has really exposed us, in fact, on how to deliver the step-down training to the teachers. So this is what we are really happy with. (TF, Jigawa)</i></p>
S3.25. p50	<p>Discussions with TFs and TDTs at endline revealed that TDP had changed their appreciation of the importance of learning in the classroom, versus copying from the blackboard, and the importance of praise – for example, by learning several different ways for pupils to clap, applauding their own or their classmates' responses – and of creating an equitable learning environment in the classroom.</p>
S3.26. p85-86	<p>As the training model has been rolled out using various sources of funding (SUBEB, UBEC and Global Partnership for Education), states have changed and, in some cases, diluted the training model. In Jigawa, the roll-out of training using SUBEB and Global Partnership for Education funds has involved a much shorter training period and has not included school support visits, which are an essential part of the TDP model. In Katsina, SUBEB funds provided by UBEC were used for training teachers using a separate UBEC cluster training model, rather than a model based on TDP. By contrast, in Zamfara, the TDP training model was used with some adjustments to remove the provision of the Trainer in the Pocket or SD card materials, to allocate trainers by Area Development Council rather than LGA, and to train some rather than all teachers in the schools (Durrani <i>et al.</i>, 2018).</p>

⁷ The 2016 formative research study was commissioned by DFID to provide an in-depth description and analysis of how TDP's teacher in-service training activities had been implemented by that stage.

S3.27.
p82

Boys' learning outcomes are on average higher than girls' in the three states (Figure 12.2). This gender learning gap has persisted between Primary 3 and Primary 6 in all three subjects. As well as being less likely to attend school at all in the three states (see Chapter **Error! Reference source not found.**), girls are disadvantaged once in school, possibly due to different household responsibilities, differences in regular attendance (although the household survey results reported in Chapter **Error! Reference source not found.** suggest that girls tend to attend school slightly more regularly than boys), or differences in the attention given by teachers to boys and girls.

Figure 12.2: Proportion of pupils reaching lowest and highest learning bands by gender



Reproduced with kind permission of: EDOREN, 2018, p. 82

S3.28.
p30

Gradual improvements in girls' enrolment are suggested to have taken place across the three states. In Katsina, a UNICEF programme is aiming to increase the enrolment and attendance of girls by supporting families with financial incentives matching the foregone income of girls not working. SBMC members and teachers emphasised the positive effect of this programme on the enrolment and attendance of girls.

S3.29.
p36

Across states and schools, teachers have to leave schools to use the toilets of neighbouring houses. Sometimes toilet facilities are shared between boys and girls, or are not maintained, so that pupils do not want to use them. In most schools, toilet facilities lack doors, which raises concerns for the safety of primary school girls. This is a major problem that affects girls' attendance and poses a challenge for female teachers.

[T]here are toilets at the back here, which are so old they need to be renovated and the female teachers do not like to go there at all [...] they go to the neighbouring houses, especially [anonymised] house, all our

	<i>morning and afternoon, female teachers go to her house. (Teacher, Katsina)</i>
S3.30. p4	The survey included interviews with the selected teachers and head teacher; lesson observations of the selected teachers (including the head teacher if he or she taught lessons); and a teacher development needs assessment (TDNA) covering English, mathematics, science, and skills in measuring pupils' progress. At baseline, eight pupils were selected randomly from those who started Primary 3 in September 2014, who were taught English, mathematics, or science by at least one of the sample teachers, and who were present in school on the day of the survey. Tests of English, mathematics, and science, and a brief interview on background characteristics, were administered to these pupils.
S3.31. p73	Pupils were assessed in English, mathematics, and science. The analysis of these test results uses item response theory, a technique which simultaneously estimates the difficulty level of each question that appeared in the test and the ability level of each pupil in the sample. By taking account of differences in difficulty between questions, it produces more meaningful results than simply calculating the percentage score on the test.

S3a is the second volume of the endline report of the Teacher Development Programme.

S3a.1. p57	Several measures depend on reporting by teachers and head teachers, and may be subject to desirability biases. This is the case for teacher motivation measures and responses about the usefulness of TDP training and materials. Teacher absence is measured both from school records and by asking teachers themselves when they were absent. Although this may affect the results, the evaluation team considers it unlikely that there would be systematic differences between treatment and control schools in the extent of this bias, and so do not think that this would bias statistics in favour of, or against, showing the effectiveness or impact of the programme.
S3a.2. p57	Measures of teacher behaviour in the classroom depend on classroom observations (see Chapter Error! Reference source not found.) and are subject to the Hawthorne effect, where being observed may affect teachers' behaviour. The results of classroom observations are perhaps best seen as teachers demonstrating the behaviour that they think is expected of them, and is not necessarily an accurate guide to how teachers teach on a day-to-day basis when not observed.
S3a.3. p78	The test was developed using the Nigerian primary syllabus and primary school textbooks to guide content appropriate for each grade. A test with approximately one-third too many items was piloted in order to examine the functioning of different test items and to select those for the final tests. The test was designed to be administered largely in Hausa, except for parts of the English test that tested understanding of spoken English, maths questions involving word problems, and some science

	questions that could not be administered without using written language (e.g. labelling a flower).
S3a.4. p78	The psychometric properties of the English, maths, and science tests were analysed within the Rasch measurement model. The more restrictive Rasch model was also applied for the baseline analysis. Items found to be less effective in assessing the traits are discussed. Items are flagged if they do not fit the Rasch model.

Non-Formal Learning Centres (NFLCs) (S4)

S4.1. p6-7	<p>1. While both ALP and ALP+Coaching (vs control) led to statistically significant improvements in the literacy, numeracy and SEL outcomes of OOS children, at a cost of £543 per child the basic ALP proved to be more cost-effective than the more expensive (£ 708 per child) but less effective ALP+Coaching:</p> <ul style="list-style-type: none"> • Literacy: The basic ALP intervention improved OOS children's literacy outcomes in 2/5 EGRA subtasks (i.e. oral reading fluency and reading comprehension). The ALP+Coaching intervention did not lead to any significant improvements in students' reading skills. • Numeracy: The ALP (vs control) led to improvements in 7/8 numeracy skills and showed to be more effective than the ALP+Coaching, which led to an improvement in 4/8 EGMA subtasks. • SEL: The basic ALP led to a decreased orientation toward the use of aggressive conflict resolution strategies, but to no other significant effects on SEL outcomes. The ALP+Coaching did not have any statistically significant effects on SEL outcomes
S4.2. p8	At a cost of £ 165 per child, adding on-site coaching to ALP led to small, negative and statistically significant effects on students' learning outcomes and mixed effects on their SEL outcomes
S4.3. p9	Our findings to date do not support the hypothesis that providing on-site coaching leads to significant improvements in the quality of learning facilitators' instructional practices
S4.4. p39	We did not observe any baseline differences between girls and boys in academic and SEL outcomes. At endline, we found that both the basic ALP and the ALP+Coaching did not have any differential impacts on girls and boys. Therefore, neither treatment led to increases or decreases in the sex-based equity gaps.

S4.5. p47	We have partnered with TEGA- Girl's Effect to conduct a qualitative study that will aim to understand the findings from this endline evaluation. Specifically, we will collect interviews with key stakeholders to understand the variety of experiences that student, teachers and coaches had in NFLC where we observed high and low levels of intervention impact.
S4.6. p19	Background Information Questionnaire: This tool collected student demographic characteristics and contact information (More in full report appendix)
S4.7. p19	EGRA is a performance-based literacy assessment designed by RTI International. EGRA has been validated in Nigeria, showing that it effectively measures comprehensive dimensions of reading competencies, such as phonemic awareness, vocabulary, letter identification, oral reading fluency, and reading comprehension among Nigerian children. We used Tangerine to collect the EGRA data.
S4.8. p19	EGMA is a performance-based numeracy assessment designed by RTI International. It measures various aspects of children's numeracy competencies, including number identification, quantity discrimination, missing number, addition, subtraction, and word problems. We used Tangerine to collect the EGMA data.
S4.9. p20	To measure children's socio-emotional skills, we used various assessments such as Children's Stories, Assessment of Children's Emotional Skills (ACES), and Moods and Feelings Questionnaire (MFQ). Children's Stories is an SEL assessment that the IRC and Creative validated in Nigeria with funding from USAID. The assessment introduces children to a series of scenarios and they are prompted to answer questions. In five ambiguous scenarios children have to imagine that another child does something that hurts them but it is unclear whether the child intended to hurt them or if the disturbance happened by accident and they are then asked to express how they would feel about the situation and interpret why the child acted in that way. In four additional scenarios, children asked to discuss how they would solve interpersonal conflicts with other children. ACES (Schultz & Izard, 1998) also presents a series of scenarios that aim to measure children' ability to recognize accurately the emotions of other people. The MFQ (Angold et al., 1995) uses 13 statements to screen depression in children and adolescents aged between 6 and 17 years.
S4.10. p22- 23	Teacher Classroom Observation (TCO): This instrument assesses learning facilitators' instructional practices in different classroom settings. To test the reliability of TCO, we adopted two reliability statistics such as Cohen's Kappa and Cronbach's Alpha. Cohen's Kappa was used to assess the inter-rater reliability of different enumerators observing the same classroom. In using Kappa with a sample of 79 classrooms observed in this study, we observed moderate to substantial levels of

	agreement among different raters ⁸ . Cronbach' α is used to assess the internal reliability of the items included in the TCO. The results show that nearly all the subscales of TCO had good internal reliability with values above .70.
--	--

Reading and Numeracy Activity (RANA) (S5)

S5.1. p4	pupils in both states and both school types display statistically significant learning gains between the beginning and the end of the school year in terms of numeracy learning outcomes. Number identification, missing number identification, addition, and word problem scores all increased by 12.4%, 11.2%, 9.4%, and 8.7%, respectively over the course of the school year
S5.2. p57	each year of project implementation has shown additional improvements in learning outcomes. These improved outcomes indicate that better results may be seen after teachers and coaches become comfortable with a project, and after pupils spend sufficient time in the program.
S5.3. p57	In addition, the improved learning outcomes indicate that some pedagogical adjustments (such as adding review periods and training teachers on classroom management techniques) may have been successful.
S5.4. p55	This year, the majority of teachers use the full lesson period, perhaps because the 2017 RANA trainings sensitized coaches and teachers to the importance of using the full class period for the lesson.
S5.5. p38	We define teacher VA as a metric that enables us to evaluate teacher performance as measured by the learning growth achieved by each individual teacher's pupils over the course of the academic year.
S5.6. p4	We find that teachers whose head and/or lead teacher attend cluster meetings have significantly higher VA scores than teachers whose respective HT and LT do not. Other characteristics of the teacher or classroom are not indicative of teacher quality such as teacher age, experience, class size, or lesson duration.

⁸ Two items showed moderate to weak between-assessor agreement: 0.56 for "Teacher helps pupils connect lesson content to their backgrounds and experiences" in Use of Teaching Methods, and 0.29 for "Teacher encourages pupils to ask questions, take guesses, express opinions during the lesson" in Pupil Participation & Well-being. We will adjust the tool and improve training at online.

S5.7. p49	it seems that the teacher's employment status and salary may be correlated with teacher performance with teachers from the top quintile earn about 4,347 Naira more and are 16 percentage points more likely to have full-time status than teachers in the bottom 20 percent
S5.8. p50	We find that teacher education is a significant predictor of teacher performance in the classroom. Moreover, we estimate that teachers with at least some postsecondary education have, on average, .23 standard deviations higher VA scores than teachers with a lower level of educational attainment. In terms of teacher characteristics and qualifications, we find that teachers with full-time status at their schools outperform their part-time or volunteer counterparts.
S5.9. p51	we find that female teachers outperform their male counterparts with similar qualifications and classrooms
S5.10. p56	teachers who follow the lesson activity steps to full fidelity have a higher value-added score in terms of their contribution to their pupils' reading ability gains
S5.11. p3	An important aspect of the RANA program is the provision of learning materials from pupil books, pencils, and classroom materials to each school. We compute the percentage of observed classrooms that have complete inputs, defined as all pupils having a pupil book and pencils, and whether pupil work is displayed on the classroom's walls. We find that 14% of all observed classrooms fulfill these inputs
S5.12. p10	These Master Trainers and SSOs trained approximately 800 educators, including P1, P2 and P3 teachers, head teachers and lead teachers. Teachers were further supported by the RANA community of practice, which includes monthly Master Trainer and SSO visits, weekly lead teacher observations, weekly school-level meetings, and monthly cluster meeting.
S5.13. p10	As part of this certification process, RANA Master Trainers and teachers received a 1-day training in Early Grade Reading Concepts in July/August 2011
S5.14. p11	RANA follows a rolling data collection whereby Master Trainers visit schools every month (three times per term,) in order to interview respondents, administer learning assessments and conduct classroom observations. There are a fewer number of head and lead teacher observations, classroom observations, and pupil observations for IQS than public school for both Katsina and Zamfara. The School Support Officers are government officials, each assigned to monitor and support between 8 and 16 schools.
S5.15. p53	Transitioning into the final year of the program, the RANA lesson plans were adapted to shift focus away from the 'I do' 'We do' 'You do' approach that is known to be more rigid and allows the teachers and pupils less freedom in the classroom and

	<p>encourages imitation. The shift in focus within the designed lesson plans is to provide the teachers with the flexibility to follow the lesson plan as designed, but with more freedom to follow different approaches to delivery as they see fit in the classroom. The second key performance indicator was previously based on following the 'I do' 'We do' 'You do' approach to fidelity. However, since this approach has become less relevant, we opted to redefine minimum teaching competencies as teachers who are observed to complete all of the lesson plan's steps correctly. This is a useful definition for the program, as we have shown it to be positively correlated with better teacher performance and is compatible with existing data from previous years of the program.</p>
S5.16. p4	<p>we calculate the percentage of girls who can read at grade-level to increase from 14.7% in the first year to 17.5% in the second, and 19.7% in the third year. This translates to a 34% increase from year 1 to year 3 in the number of girls who can read at least 31 words per minute correctly.</p>
S5.17. p11	<p>RANA follows a rolling data collection whereby Master Trainers visit schools every month (three times per term,) in order to interview respondents, administer learning assessments and conduct classroom observations.</p>
S5.18. p11	<p>From February to May of 2016 the RANA team trained 57 enumerators, comprised of 40 Master Trainers and 17 School Support Officers. The School Support Officers are government officials, each assigned to monitor and support between 8 and 16 schools. The Master Trainers were recruited by the RANA team from local Colleges of Education to assist with monitoring and provide pedagogical support to 3 to 6 schools. During the 5-day training all participants were provided with Android tablets and shown how to use the Open Data Kit (ODK), Tangerine and Random Number Generator applications. Enumerators then had an opportunity to practice observing classes and assessing pupils at schools near the training site. Refresher trainings were held in October 2016, January 2017 and October 2017.</p>
S5.19. p38	<p>In this report, we introduce and implement the concept of estimating a teacher value-added model with several aims in mind. First, this exercise provides reading and numeracy interventions such as RANA the ability to determine heterogeneity in the performance of the schools implementing the intervention itself. This means that the intervention will be able to distinguish between schools and teachers in the program based on their pupils' learning outcomes. The intervention and administrators will be able to provide targeted and, to a certain extent, customized support for specific schools. Second, the project will be better equipped to identify teachers and schools who need additional support as well as an opportunity for the project to learn from high performing teachers and schools and apply these lessons learned in other schools. Lastly, implementation of a teacher value-added model in RANA allows us to investigate the correlates and/or determinants of pupil learning growth that better informs project decision making from a technical and operational standpoint.</p>

S5.20. p42	From an econometric standpoint, it is important to note that assessments in general, as well as EGR/MA pre- and post-tests, are measured with some degree of error.
S5.21. p42	To account for measurement error in the pretest we employ an error-in-variables (EiV) regression model that accounts for the reliability of each assessment in the estimation of each of the regression coefficients.
S5.22. p57	we reiterate the importance of the RANA monitoring and evaluation system that was effectively implemented in the RANA context and now transferred to the state authority of Zamfara as part of their state's larger educational system. We show that the RANA system is straightforward to implement as it relies on standard M&E tools that most early grade reading projects implement in similar contexts.
S5.23. p57	However, we emphasize the importance of pupil-level longitudinal tracking within year, rather than employing repeated cross-sections, that enable us to make teacher performance determinations and ascertain heterogeneity in the program's overall performance. These are informative metrics that are typically missing from early grade reading interventions that use similar M&E tools.

Northern Education Initiative Plus (S6)

S6.1. p8	there remain several key factors limiting the program's impact in terms of learning outcomes. Key among these factors, which were signalled in the baseline report and have since been further investigated through project monitoring mechanisms, are high rates of absenteeism among both teachers and students; insufficient time on task more generally; and low capacity and motivation of the education officials who are tasked to serve as coaches to teachers. Additional measures to address these factors, as well as reinforcement of EGR training for teachers, could be expected to improve learning outcomes further by end line.
S6.2. p95	Moreover, three fundamental constraints identified at baseline were flagged as posing challenges to the program's ability to effect significant changes in learning outcomes – even with substantial investment in inputs to improve the quality of teaching and the frequency and quality of school support: <ul style="list-style-type: none"> • Learners' preparedness for school, as indicated by frequent absences from school, lack of adequate nutrition at home, and related issues (e.g., low parental awareness of, and support for, formal schooling). • Teachers' frequent absence from the classroom (as noted in prior EGRA reports).

	<ul style="list-style-type: none"> School supervisors' infrequent visits to classrooms, deriving from a constellation of obstacles, and resulting in inadequate support of teachers of reading.
S6.3. p95	<p>TLMs and regular teacher training were designed to address these issues, over time, and the present mid-line report presents evidence of modest impact from these inputs. For example, as a result of training provided by the Initiative in 2017, teachers are spending more time teaching reading skills and doing so more comprehensively. Not surprisingly, small but statistically significant correlations are found between the teacher practices index and reading scores as well as between lesson duration and reading scores. We expect that greater learning improvements would have resulted from additional teacher training in advance of the mid-line EGRA. Unfortunately, budget constraints forced the cancellation of teacher training that was planned for this period. The inability of the Initiative to provide this additional training can be assumed to have inhibited teachers' competency gains and to have negatively affected teacher motivation and attitudes towards teaching. EGRA mid-line data reflect this limited improvement in instructional quality, which has hampered the Initiative's ability to impact learners' performance more significantly.</p>
S6.4. p96	<p>with the above-mentioned findings on teachers' attitudes and knowledge of EGR, indicate the need to take the following actions: increase training, regular coaching and mentoring, and reinforce other teacher professional development support; and 2) further improve and focus training <i>content</i> to strengthen development of learners' skills in key areas. Specifically, the Initiative will -provide additional, practice-based teachers' training content and activities. New content focuses on deepening teachers' emerging EGR knowledge and skills, especially toward reinforcing decoding, fluency and listening comprehension, as well as the effective use of formative assessment.⁶ This approach will include new materials as well as instructional techniques (e.g., pairing learners) to better support learning. -improved supervision and support to improve instructional practice at the school level. This latter effort includes reinforced coaching and mentoring support (i.e., via training and 'coaching of coaches') and new, structured teacher professional development content to better support the development of strong, effective communities of practice</p>
S6.5. p93	<p>In Hausa for both P2 and P3, mean scores increased for all subtests, and zero scores decreased. The most pronounced improvements in P2 were in syllable identification and the writing of letters in dictated words. These improvements suggest that a portion of teachers have improved their effectiveness in facilitating pupils' practice in decoding and encoding.</p>
S6.6. p97	<p>Among the additional factors that may be assumed to have contributed to the limited reading performance gains are challenges with access to Mu Karanta! and Let's Read! Teacher's Guides and Pupil's Books in target LGEAs. Timely availability of these materials has been complicated by a number of factors, including:</p> <ul style="list-style-type: none"> Inadequate SUBEB capacity to procure, store and distribute materials to LGEAs;

	<ul style="list-style-type: none"> • Lack of LGEA resources to deliver materials to schools; • Poor communications between SUBEB, LGEA and school officials regarding distribution protocols and schedules
S6.7. p94	<p>Even where the incidence of certain practices has increased (e.g., the number of teachers asking learners to read at home), the level of application of best practices observed among teachers is still low overall (60% or less in many cases). These persistent, low levels of application of best instructional practices in reading are evidently <i>too low</i> to adequately support a substantial increase in learners' reading performance.</p>
S6.8. p94	<p>It is also essential to emphasize that Initiative efforts to improve reading performance are taking place in an environment in which systemic governance remains weak. Gaps in policy, education planning, lack of follow through on fundamental decisions (e.g., budget allocation and release of funds for TLM provision), and inconsistent follow through by state actors on TPD and school support plans, among other accountability and capacity-related issues, all point to system weaknesses that the Initiative has only been able to address partially. These, and other, related contextual factors were highlighted in the findings of the baseline EGRA report</p>
S6.9. p95	<p>Additional highlighted factors included:</p> <ul style="list-style-type: none"> • Exceedingly low level of foundational skills among learners; • Inadequate training and significant knowledge gaps in teachers' knowledge of EGR concepts, principles and practices; • Inadequate and/or infrequent school supervision or pedagogical support; and • Lack of reading resources to support teaching and learning. <p>Moreover, three fundamental constraints identified at baseline were flagged as posing challenges to the program's ability to effect significant changes in learning outcomes – even with substantial investment in inputs to improve the quality of teaching and the frequency and quality of school support:</p> <ul style="list-style-type: none"> • Learners' preparedness for school, as indicated by frequent absences from school, lack of adequate nutrition at home, and related issues (e.g., low parental awareness of, and support for, formal schooling). • Teachers' frequent absence from the classroom (as noted in prior EGRA reports). • School supervisors' infrequent visits to classrooms, deriving from a constellation of obstacles, and resulting in inadequate support of teachers of reading.
S6.10. p96	<p>Among the additional factors that may be assumed to have contributed to the limited reading performance gains are challenges with access to Mu Karanta! and Let's Read! Teacher's Guides and Pupil's Books in target LGEAs. Timely availability of these materials has been complicated by a number of factors, including:</p>

	<ul style="list-style-type: none"> • Inadequate SUBEB capacity to procure, store and distribute materials to LGEAs; • Lack of LGEA resources to deliver materials to schools; • Poor communications between SUBEB, LGEA and school officials regarding distribution protocols and schedules; • Inadequate training of personnel tasked with managing and carrying out the procurement, storage and/or distribution of the materials; and • Inadequate state government resources, delayed resource allocation/release, and/or poor planning and follow through (e.g., due to accountability issues).
S6.11. p33	<p>Head teachers were also asked about the frequency of monitoring and supervision of Hausa Primary 2 teachers by decentralized officers of SUBEB (Area Education Officers and School Support Officers). At baseline, over two-thirds of the head teachers interviewed in Bauchi and about 40% of head teachers interviewed in Sokoto responded that such supervisory visits never occurred. At midline, the percent of head teachers in Bauchi responding never has dropped drastically to about 6%, with corresponding increases in responses of weekly and monthly supervisory visits. In Sokoto, fewer head teachers responded never at midline as well, though the decrease in never responses was not as significant.</p>
S6.12. p95	<p>Additional highlighted factors included: ... Inadequate and/or infrequent school supervision or pedagogical support</p>
S6.13. p96	<p>These findings, combined with the above-mentioned findings on teachers' attitudes and knowledge of EGR, indicate the need to take the following actions: Specifically, the Initiative will:</p> <ol style="list-style-type: none"> 1) increase training, regular coaching and mentoring, and reinforce other teacher professional development support; and 2) improved supervision and support to improve instructional practice at the school level. <p>This latter effort includes reinforced coaching and mentoring support (i.e., via training and 'coaching of coaches') and new, structured teacher professional development content to better support the development of strong, effective communities of practice.</p>
S6.14. p4	<p>Consistent with the system strengthening agenda of the Initiative, the assessors were drawn from state education institutions, in particular the State Universal Basic Education Board (SUBEB) and the LGEAs. Training for the midline data collection was held from June 25-30, 2018 in Bauchi and July 16-20, 2018 in Sokoto.</p>
S6.15. p94	<p>It is also essential to emphasize that Initiative efforts to improve reading performance are taking place in an environment in which systemic governance remains weak. Gaps in policy, education planning, lack of follow through on fundamental decisions</p>

	(e.g., budget allocation and release of funds for TLM provision), and inconsistent follow through by state actors on TPD and school support plans, among other accountability and capacity-related issues, all point to system weaknesses that the Initiative has only been able to address partially.
S6.16. p97	As explained above, continued and wide gaps in states' ability to carry out fundamental, system-level functions (e.g., performance monitoring) threaten to undermine both the performance gains made to date, as well as the prospects for lasting change. In this vein, the Local Education Monitoring Approach (LEMA) has demonstrated considerable promise as a simple, practical, and informative mechanism for which buy-in has been encouraging in both states. Additional system strengthening actions may be taken (with support and in coordination with USAID) at the highest levels of decision making in each state, using the evidence gathered to support substantive changes in governance, as described here.
S6.17. p3	The Initiative's Early Grade Reading Assessment (EGRA) and related assessments in formal schools are employing a cross-sectional methodology; specifically, baseline measurements of the reading outcomes of Grade two and Grade three learners in Y1 (before any EGR intervention) are being compared to those of subsequent cohorts of learners taken at the same time of the school year in Y3 and Y5.
S6.18. p21	In spring 2017, NEI Plus conducted a four-day workshop to review the existing EGRA instruments (used for the baseline EGRA administration) and develop additional versions that were comparable for administration during midline and endline. These versions were subsequently pilot tested using common-persons equating, as recommended in the EGRA Toolkit.
S6.19. p22	An additional limitation derives from that fact that time, budgetary, and logistical constraints required the training of assessors and the data collection to be conducted separately in the two states. Faced with this situation, measures were taken to ensure the maximum possible harmonization of the training and the data collection procedure between the two states, and indeed the same experienced head trainers led the training in both states. Nonetheless, the practical separation of the Bauchi and Sokoto EGRA baseline exercises meant that IRR tests were not conducted <i>across</i> the two states. As a result, any comparison of EGRA and of the classroom observation results between states should be undertaken with caution and considered only as potentially indicative.
S6.20. p103	Following the tool development workshop, a pilot study was implemented to further ensure comparability between the different versions. The pilot was implemented following a common-persons equating design, as specified in the <i>EGRA Toolkit</i> , in which each sampled student received two administrations of the Hausa and English EGRA instruments: the baseline version (Version A) and a new version (either Version B or C). As specified in the <i>Toolkit</i> , the sample was designed to lessen the testing burden on students with each student receiving

	two versions, but not all three, of the EGRA instruments. During administration, each student was randomly assigned to receive either Version B or C.
--	---

The effects of female teachers on girls enrolment (S7)

S7.1. p3	Female teachers generally have a positive impact on girl pupils, both in Nigeria and globally (Action Aid 2012; Forum for African Women Educationalists 2010; Paulson 2008). The impacts include greater enrollment and retention rates of female pupils, increased exposure to female role models, more girl-friendly school environments, girls' increased ability to articulate barriers to their education, and lower rates of girls dropping out of school (Tao 2014).
-------------	---

Teaching at the Right Level (TaRL) (S8)

S8.1. p3	The TaRL remedial learning programme builds on the LEARNigeria citizen-led household assessment. Results from the 2017 LEARNigeria survey (focusing on TaRL intervention states) are highlighted.
-------------	---

ENGINE II Learning needs assessment (S9)

S9.1. p4	For the out-of-school girls, 35% of the girls who are currently were able to respond to questions from JSS1 to SSS3. On the other hand, girls who are in continuing education, 60% of them could not go beyond primary level questions. This poses a problem for the programme and there is another need for the programme to rethink about its strategy. Continuing education is meant for girls who have dropped out of school and are willing to continue their education or undertake their school leaving exams. Whereas, basic classes are for girls who have no or little knowledge in literacy and numeracy. From the data, 35% of the OSGs who are currently in the basic classes should either be in post-basic or in continuing education, whereas 60% of the girls who are in continuing education need to be in basic classes.
-------------	---

Research on female teachers (S10)

S10.1. piii	The presence of a single female teacher has a positive and statistically significant effect on the percentage of girls in total enrolment. Moreover, schools with at least one female teacher are associated with a 1.5 percentage point increase in girls' share of total enrolment.
----------------	---

S10.2. piv	each 10% increase in the number of female teachers increases girls' enrolment by 0.44 percentage points. This means that, all else being equal, both the mere presence of female teachers, and the 'dosage' thereof, increases girls' enrolment by a statistically significant margin. These findings look somewhat different across rural and urban contexts
S10.3. pvi	The presence of female head teachers in rural schools is not as impactful as having female teachers in the classroom. We find that schools with male or female head teachers provide a similar likelihood of influencing girls' enrolment. This was supported by the qualitative evidence, which found that engaged and supportive head teachers are found across genders, with no key differences between male/female or rural/urban. Among urban schools, however, female head teachers play a more substantial role in improving girls' enrolment and promotion rates than in rural schools.
S10.4. pv-vii	From our quasi-experimental analysis, we found that female teachers are more effective in improving girls' learning outcomes than male teachers. We estimate that female teachers increase girls' reading gains by about 25% of a standard deviation, and 15% of a standard deviation in math gains, relative to similar girls taught by a male teacher. This could mean that girls are more likely to respond positively to instruction under a female teacher, and/or that female teachers are more effective in teaching to girls than their male counterparts.
S10.5. pvii	we find that boys' reading gains are also positively influenced by female teachers, such that boys exhibit an additional 12.9% of a standard deviation gain over that of similar boys taught by a male teacher. This finding supports the idea that female teachers are more effective in teaching.