A Quasi-Experimental Study of the Classroom Practices of English Language Teachers and the English Language Proficiency of Students, in Primary and Secondary Schools in Bangladesh.

Research Report

Authors

Tom Power, Robert McCormick and Elsbeth Asbeek-Brusse.

A Quasi-Experimental Study of the Classroom Practices of English Language Teachers and the English Language Proficiency of Students, in Primary and Secondary Schools in Bangladesh.

Research Report

English in Action (EIA) (2017) A Quasi-Experimental Study of the Classroom Practices of English Language Teachers and the English Language Proficiency of Students, in Primary and Secondary Schools in Bangladesh. Dhaka, Bangladesh: EIA.

Authors

Tom Power, Robert McCormick and Elsbeth Asbeek-Brusse.

Acknowledgements

Executive Summary

Background

English in Action (EIA) is an English language teacher development project based in Bangladesh that was intended to run from 2008 to 2017, but which was extended at the request of the Government of Bangladesh, with additional funding from UKAID, for a further year to 2018. By the time of the design of this study (2014-2015) EIA was drawing to the end of upscaling (phase III, 2011-2014) and entering institutionalisation and sustainability (phase IV, 2014-17, extended 2018). Successive prior studies had indicated substantial success in improving both teachers' classroom practices and student learning outcomes, over the pre-project baseline (e.g. EIA 2011, 2012). The 2014 Annual Review of EIA recommended that in the final phase, EIA should explore whether it would be possible to carry out a study that compared a 'counterfactual' or control-group of teachers and students, to the 'EIA' or treatment schools: i.e. a Randomised Control Trial or Quasi-Experimental study. A proposal for a Quasi-Experimental study was developed in collaboration with DFID's South Asia Research Hub (SARH), which also provided the additional funding necessary to implement such a study.

The teachers and students who were the subject of this study, were the fourth cohort to participate in English in Action (together with teachers from 'control' schools, in the same Upazilas). This fourth EIA cohort included Schools, Teachers and Students from approximately 200 Upazilas (of approximately 500 in total) across Bangladesh, including some of the most disadvantaged areas (with reference to UNICEF deprivation index), such as Char, Hoar and Monga districts.

Teachers took part in a school-based teacher development Programme, learning communicative language teaching approaches through carrying out new classroom activities, guided by teacher development videos that showed teachers, students and schools similar to those across the country. Teachers also had classroom audio resources for use with students. All digital materials were available offline, on teachers own mobile phones, so there is no dilution of the Programmes core messages about teaching and learning, by some intermediary coming between the teacher and the materials. Teachers were supported through these activities, by other teachers in their schools, by their head teachers and by local education officers. Some teachers from each area were also given additional support and guidance from divisional EIA staff, to act as *Teacher Facilitators*, helping teachers work through activities and share their experiences at local cluster meetings. Whereas previous cohorts of teachers had attended eight local teacher development meetings over their participation in the project, for Cohort Four, this was reduced to four meetings, with a greater emphasis being placed on support in school by head teachers, as well as support from local education officers. This change was part of the move towards institutionalisation and sustainability of project activities within and through government systems and local officers.

The purpose of this study was both to provide the evaluation evidence required for the final phase of the EIA project and to contribute to the international body of research evidence on effective practices in teacher development in low-to-middle income country contexts.

Research Methodology

The study used a 'quasi-experimental' design[1], in which schools in the study were randomly assigned to one of two groups: control schools (which would not take part in EIA in cohort 4[2]), or treatment schools (which would take part in EIA in cohort 4). A pre-test was planned to take place before the treatment began (i.e. before the treatment schools began taking part in EIA) and a post-test was planned to take place as close to the end of the treatment as possible (i.e. around or just after the time of the fourth of four teacher development meetings). The study includes two measures, each with a different focus and method of data-gathering:

1. Students' communicative competence in English language, assessed through Trinity College London's Graded Examinations in Spoken English (GESE). These assessments were carried out through diagnostic interview, by Trinity College accredited assessors, from South Asia.
2. Classroom Practices, assessed through timed-observations of patterns of teacher and student talk during English lessons, as indicators of communicative language teaching practices. These observations were carried out by post-graduate fieldworkers from The Institute of Education and Research (IER) at Dhaka University (DU).

For English Language Competence (ELC), some 1,802 individual students were assessed at both pre- and post-test (3,484 students were assessed at pre-test, but the findings are based only upon individual students for whom we have both pre- and post-test data), distributed across primary and secondary and control and treatment groups[3]. For classroom practice, some 163 individual teachers' lessons were observed pre- and post-test (242 teachers at pre-test, but finding based only upon individual teachers for whom we have both pre- and post-test data), from the same sample of schools as for the ELC measure.

The intention of the design was for the study to take place in one school year, with all School-Based Teacher Development (SBTD) activities taking place between March – September 2015, with pre- and post- test just before and after these dates. However, the nationwide security disruption in early 2015 forced the SBTD activities to be pushed back, straddling two school years. The post-test could not be delayed beyond May 2016 as funding for academic analysis and reporting was due to finish July 2016, with project close-out by March 2017. The final design, agreed with DFID and SARH (May 2015), put pre-test put back to May 2015, post-test May 2016.

The SBTD Programme was then unable to begin until August 2015, with teachers completing in September-October 2016 (*see 2.5   Timeline*). At the time of the post-test (May 2016), teachers were only half-way through the EIA school-based teacher development Programmes (having completed two of four teacher development meetings). But in reality, the only effective 'teaching time' available for the introduction of new classroom activities was limited to around two-to-three months, from February/March to April/May 2016. This *partial* treatment provided only limited time

---

[1] This is very similar to a Randomised Control Trial, but adapted to contexts where individuals cannot be randomly selected from the entire population, but must be selected from specific sub-sets of the population (i.e. in this instance, the Upazilas where the project had government authority to be working in phase IV).
[2] For ethical reasons, the 'control' schools were scheduled to take part in EIA in the year following the study, so that they were not disadvantaged.
[3] See Table 5: Actual sample for student ELC assessments (GESE)

for teachers to integrate EIA classroom activities into their regular practice and an unrealistically short period for these changes to translate into measurable improvements in learning outcomes.

The study design allowed for up to 33% student attrition (Table 2). However, attrition was over 48% (Table 5). 89% of attrition was due to students being absent at post-test and 11% of attrition was due to schools being withdrawn from the study (Table 7).

Key Findings

Classroom Practices

The pre-test findings were broadly in-line with those suggested by the original baselines (EIA, 2009), but the two key indicators of classroom practice for the Logical Framework (logframe) were both somewhat higher than anticipated. Pre-test student talk-times were found in the range of 16-24% of lesson time[4], with 74-88% of student talk in English[5], although these pre-test levels of student talk and student talk in English at pre-test were not associated with correspondingly higher levels of other communicative practices, such as increased student-student talk, or improved ELC.

The project teams' interpretation of this, is that through prior exposure of education officers and teachers to EIA in many of the experimental Upazilas (though not the experimental schools directly) and through institutionalisation of EIA materials, approaches and employment of EIA Teacher Facilitators in national government Programmes of *Subject-Based Training* and *Curriculum Dissemination Training*, the general principles that *it is good for students to talk* and to *talk in the target language*, appear to have become widely accepted across the school system, beyond those schools that have directly participated in EIA. However, pre-test data also suggests that whilst these broad principles may have been put into practice, there was still a predominance of teacher-talk presenting, with little student-student talk, and little evidence of a wider shift in improved English Language Competence.

The key findings of the study show experimental effects on classroom practice, with statistically significant difference-in-differences between control and treatment and effect sizes[6] in small to medium range. These effects broadly show a shift away from traditional classroom practices towards more communicative approaches.

1. All **teachers spent less time 'presenting'** to students, as a result of the EIA treatment[7].
    a. In treatment schools, presenting dropped from 49% of teacher talk to 43% (this decrease was statistically significant), whilst in control schools there was no significant change.
    b. The difference-in-differences, i.e. between treatment and control schools of a 9% point reduction in teacher talk presenting, was statistically significant and the effect size (r=0.2) was small-to-medium.
2. In primary lessons, **student-student talk (i.e. pair or group work) increased** fourfold, as a result of the EIA treatment[8].

---

[4] See Table 19: Student Talk time as a percentage of lesson time, summary experimental results
[5] See Table 22: Student talk in English as percentage of student talk-time, summary experimental result
[6] See Appendix 2, for explanation of effect size interpretation, comparison and calculation.
[7] See Table 36 Teacher talk presenting as a percentage of all teacher talk, all classrooms (N=163)
[8] See Table 28: Student-Student talk as percentage of all student talk, Primary (N=79)

    a. In primary treatment schools, student-student talk increased dramatically from 4% to 16% of all student talk (this increase was highly statistically significant), whilst in control schools, there was no significant change.

    b. The difference-in-differences of a 15% point increase in student-student talk, was highly statistically significant and of a medium effect size (r=0.3)

3. In secondary lessons, **teachers used more spoken English**, as a result of the EIA treatment[9].

    a. In secondary treatment schools, teachers' use of English increased substantially from 64% to 86% (this increase was very highly statistically significant), whilst in control schools, there was no significant change.

    b. The difference-in-differences of a 16% point increase in teachers' use of spoken English was statistically significant and of small-to-medium effect size (r=0.2)

4. In secondary lessons, **students used more spoken English**, as a result of the EIA treatment[10].

    a. In secondary treatment schools, students use of English increased from 77% to 85% (this increase was statistically significant), whilst in control schools, there was no significant change.

    b. The difference-in-differences of a 15% point increase in students' use of spoken English was statistically significant and of small-to-medium effect size (r=0.2)

In addition to these statistically significant experimental effects, there were two other indicators suggesting a shift towards more communicative practices in treatment schools, but where the difference-in-difference was not statistically significant:

5. **Teachers spent less time talking**, as a result of the EIA treatment.

    a. In all treatment schools (primary and secondary combined)[11] there was a reduction in teachers talk time from 53% to 49%, which was statistically significant. There was no statistically significant change in control schools.

6. **Teachers spent more time organising student activity**, as a result of EIA treatment.

    a. In all treatment schools (primary and secondary combined)[12] teachers spent more time organising student activity from 20% to 27% of teacher talk, which was highly statistically significant. There was no statistically significant change in control schools.

English Language Competence

The pre-test findings for English Language Competence were broadly in-line with those of the original EIA baselines (2010, reported in EIA, 2011).

In terms of the project logframe indicators, EIA school post-test findings showed large and highly statistically significant improvements in relation to lower GESE grades[13], with 19% more students achieving GESE grade 1 or above in primary (rising from 40% to 59%) and 12% more students achieving GESE grade 2 or above in secondary (rising from 40% to 42%). There were also highly statistically improvements at the higher GESE grades, but these were relatively small (2% increases in both primary students achieving at or above GESE 2 and secondary students achieving at or above GESE 3).

---

[9] See Table 35 Teachers Talk in English, Secondary (N=84)

[10] See Table 25: Students talk in English, Secondary (N= 84)

[11] See Table 30: Teachers talk as percentage of lesson time, all classrooms (N=163)

[12] See Table 39: Teachers organising activity as a percentage of total teacher talk, all classrooms (N=163)

[13] See Table 13: Logframe Results, Improvements over pre-test in EIA schools (GESE)

Improvements in English Language Competence (ELC) were seen for across all students, in both control and treatment groups[14]. In treatment schools, the differences between pre- and post-test ELC were larger than in control schools (rising by 0.3 GESE grades in treatment, compared to 0.2 grades in control) but the differences between control and treatment were not statistically significant. Whilst for the study population as a whole (all students, primary and secondary) the difference-in-differences between treatment and control was not statistically significant, when the data was disaggregated by primary/secondary, gender or location, several statistically significant experimental effects were found:

1. In treatment schools, **Male students (primary and secondary, all locations) improved their English Language Competence** more than in control schools[15].
   a. In treatment schools, male students increased 0.3 GESE grades, from 1.0 to 1.3 (this was highly statistically significant). In control schools, male students also showed highly significant improvement, but only of 0.1 GESE grades (from 1.3 to 1.4).
   b. The difference-in-differences between treatment and control schools of 0.2 GESE grades was highly statistically significant, with a small effect size (r=0.1).
2. In treatment schools, **urban students (primary and secondary, boys and girls) improved their English Language Competence** more than in control schools[16]
   a. In Urban treatment schools, students increased 0.3 GESE grades, from 1.8 to 2.1 (this was highly statistically significant), where there was no significant difference in control schools.
   b. The difference-in-differences between treatment and control schools was large, almost half (0.4) a GESE grade and was highly statistically significant, with a small-to-medium effect size (r=0.2).
3. In secondary treatment schools, **rural students (secondary boys and girls) improved their English Language Competence** more than in control schools[17].
   a. In secondary rural treatment schools, students increased 0.5 GESE grades, from 1.24 to 1.74 (this improvement was highly statistically significant). In control schools, the increase was also significant, but was less than a third of this, at 0.14 GESE grades, from 1.56 to 1.7.
   b. The difference-in-differences between treatment and control schools was large, at over a third (0.36) of a GESE grade and was highly statistically significant, with a small-to-medium effect size (r=0.2).

Summary

Despite limited available teaching time (February-May 2016) to put EIA activities into practice, the study shows a number of statistically significant experimental effects on classroom practice (less teacher 'presentation', more spoken English, more student-student talk) indicating the beginnings of a general shift towards the adoption of more communicative approaches to English Language Teaching. Over the period of the study, students' English Language Competence increased in both control and treatment schools, though the increases were larger in treatment schools. However, statistically significant experimental effects were only seen for certain sub-groups at this stage.

---

[14] See Table 15: Students Proficiency (GESE), summary experimental results
[15] See Table 44: All Male students' GESE (N= 748)
[16] See Table 52: All Urban students GESE (N=203)
[17] See
Table 60: Secondary rural GESE (N= 458)

1.   Introduction

1.1   About English in Action (EIA) and this study

English in Action (EIA) is an English language teacher development project based in Bangladesh that was intended to run from 2008 to 2017. Following successful Development, Pilot and Upscaling phases (phases I, II and III), the project entered its fourth and final Institutionalisation phase, in March 2014. EIA was due to conclude in March 2017, but in 2016, received an additional funding to extend for a further year to March 2018.

By the time of this study, English in Action (EIA) had been operating for a number of years. Prior studies had shown it was very successful in improving teaching practice and learning outcomes. However, these studies had not adopted a quasi-experimental approach, which was recommended in the Annual Review (2014):

> In general, a limitation of EIA is the lack of a counterfactual or a quasi-experimental design. This could be described as a 'legacy' issue – when EIA was being developed, there was less focus on the need for a counterfactual. … [and this] makes it difficult to fully attribute pupils' English language learning gains to the EIA intervention. (Annual Review EIA, 2014, p. 4)

A Theory and evidence-based approach to Teacher Development

EIA took a theory-based approach to Teacher Development, working from a developing evidence base of 'known' critical elements for improving the quality of teaching and learning through teacher development. The underlying theory-of-change for the EIA school component was that students learn more when they are taught by competent and effective teachers, and that teachers become more skilled (in terms of practice) and knowledgeable (both in terms of pedagogy and subject-knowledge) through appropriate teacher development. The most effective teacher development focuses upon classroom behaviour for both the teacher and the students, supported by learning materials for use with students (Boissiere, 2004). There are complex re-enforcements, between new teacher and student behaviours, improved students' satisfaction and interest in schooling (reflected in attendance, enrolment and achievement), and improved teacher motivations.

In terms of what constitutes effective teacher development, a recent DFID funded rigorous literature review (Westbrook *et al.*, 2013) identified four key findings:

1. Professional development aligned with teacher's needs, including focus on classroom practices (for example, through lesson modelling), with follow-up support and monitoring (providing opportunities for feedback and reflection on practice).

2. Teacher Peer Support (formal and informal peer support in clusters or schools; focused on introduction of new classroom practices; joint observations, lesson planning & resource sharing).

3. Head Teacher Support (awareness of, and support for new methods of teaching, from Head Teacher and wider school community).

4. Alignment with curriculum and assessment (school and external).

1

EIA illustrated how all these key elements could be implemented coherently in a large-scale teacher development (TD) Programme:

1. *Follow-up support and monitoring*: lesson modelling was provided through AV materials on micro SD-cards (on low-cost mobile phones); these were also used in peer-led cluster meeting activities, providing constructive feedback and discussion on practice. Monitoring took place through evaluation of these meetings, support and use questionnaires, and classroom visits.

2. *Peer support*: Teachers worked in pairs from each school. This was strengthened by regular periodic meetings with other local teachers, over a period of a year to review, reflect and introduce gradually new elements to their teaching practice.

3. *Head Teacher support*: Head Teachers were made aware of new methods of teaching being introduced and their active support was developed through Head Teacher meetings and activities. Primary Head Teachers also participated as Teachers of English, as appropriate.

4. *Alignment with curriculum and assessment*. The EIA Programme aligned closely with the curriculum (referring throughout to the English for Today textbooks, in examples of practice). Anecdotal evidence suggests many teachers and head teachers attributed improved English exam results to their participation in EIA.

In addition to the four elements identified in the international literature, EIA also identified an innovative role for offline audio-visual (AV) and print materials, which enabled both scale and impact, without reliance on national or international English Language Teaching (ELT) experts. EIA positioned mobile technology and offline media as a fifth 'critical element' for Teacher Development at scale.

5. *Offline Audio-visual (AV) materials and enabling technology*. Curriculum expertise and lesson modelling was provided through the AV materials, used in teacher and facilitator development activities carried out with peer support at local meetings and in school. This so called 'tutor in the pocket' countered the attenuation of outcomes usually associated with cascade models by providing authentic models of classroom practice directly to the teacher.

It is not possible at this stage in the research evidence to isolate the contribution of any one element, but rather we see the elements working together in an integral manner to produce a functioning system. For example, the AV materials (5) on their own are of little value, as without the other elements of support (1-4) they are unlikely to be used effectively. Equally, whilst other elements (1-4) are of general benefit, their impact is maximised by the AV materials (5).

The EIA Teacher Development Programme in practice

The Primary and Secondary School-Based Teacher Development (SBTD) Programmes were professional development Programmes in which the key site of learning was each teacher's own school and classroom. Supported by authentic video materials, teachers engaged in new classroom activities designed to develop key pedagogic skills and strategies. The video resources included sequences showing real Bangladeshi teachers using these techniques in their own classrooms as well as material that enabled teachers to reflect on their own practice individually, together with a partner teacher in their own school, and with other teachers at regular, locally-organised meetings (cluster meetings). These meetings were led by specially recruited and trained practicing teachers,

called Teacher Facilitators. The EIA approach was, therefore, very different from traditional training based on one-off, out-of-school training events, which often follow the 'cascade' principle (where successive levels of trainers attempt to 'train' the next level down, without time or contexts in which to develop knowledge and practice effectively). According to EIA research studies, the attenuation of outcomes usually associated with cascade models did not appear to be a feature of the project.

The EIA approach to Research Monitoring and Evaluation

A systematic approach to research, monitoring and evaluation (RME) was identified as a major factor in ensuring EIA's success throughout its lifetime. The Quality Assurance strategy gathered data from cluster meetings and workshops, from individual teachers and through classroom observations. This monitoring provided firm evidence that EIA training and resources were valued by teachers (e.g. typically more than 95% of teachers agreed or strongly agreed with a set of positive statements about the cluster meeting; 80% of secondary and 90% of primary teachers used English rather than Bangla for more than half of their classroom talk), but also enabled remedial steps to be taken where necessary, by the EIA Programme team.

At the level of research and evaluation, in Phase II and Phase III large-scale studies were undertaken into teachers' and students' perceptions, teachers' and students' English Language Competence and teachers' classroom practice. Phase III studies were referenced to the baselines established in the initial studies carried out in Phase II. As well as providing evidence of Programme outcomes, these studies informed the design and refinement of the Programme itself as it developed, and enabled EIA to contribute to the wider knowledge base of education and international development.

Relationship between the study and National Teacher Development Programmes

Experience of working with the Government of Bangladesh (GoB) indicated that existing capacity was insufficient to secure both (a) a strong impact on the students or teachers and (b) the full institutionalisation of EIA approaches within the national education infrastructure beyond the project end. The project therefore formulated the Phase IV action plan whereby instead of rolling out the Programme through DPE's Subject Based Training for primary teachers, and the Total Quality Initiative (TQI) In- Service Teacher Development Programme for secondary teachers (the existing GoB teacher development Programmes), a partially institutionalised mode of delivery would assure the quality of the implementation of the Programme and therefore the learning outcomes, in 2015. This partially institutionalised approach saw existing GoB Programmes of one-off teacher training events in primary and secondary supplemented by a period of school-based teacher development provided through EIA. This combination, described below, is the approach examined in this study.

Primary: Subject Based Training through PEDP III

The Department for Primary Education (DPE) ran 6-day one-off `Subject Based Training' (SBT), which was intended to reach all primary teachers during the period 2013-2016. This was essentially a cascade, with 170 'Master Trainers' cascading training through 840 'Teacher Trainers', to reach 60,000 Teachers.

As SBT trained only one teacher per school, EIA selected (for both control and treatment groups) primary schools in which one teacher has already taken part in SBT. This meant there should have been no further engagement with SBT by teachers from the sample schools, over the life of the study.

Secondary: In-Service Training through Total Quality Initiative (TQI) II
The Department for Secondary and Higher Education (DESHE) did not offer sector-wide professional development Programmes at the time of the study. In early 2015, the main secondary Programme 'Total Quality Initiative' (TQI-II) refocused its implementation plans for 2015 and 2016 to focus exclusively on Higher Secondary (Classes 9 & 10). EIA did not work with these classes, so there was no opportunity for contamination with the study.

1.2   Prior Research Findings

Research findings in both Phase II (with 600 teachers) and Phase III (with 4,000 teachers in its first cohort) had been extremely encouraging. In the crucial area of student learning outcomes, English Language Competence (assessed by Trinity College London against its GESE scale) showed statistically significant improvements over the 2010 baseline for both secondary and primary students, with the Phase III primary students outperforming their counterparts in Phase II despite the fact that their numbers were far greater. The classroom practice studies were able to show substantial improvements in key aspects of pedagogy, in comparison with a baseline study in which the teacher-dominated lessons showed very little evidence of interaction. In Phase III, for example, 27% (primary) and 24% (secondary) of the talking in lessons was done by students, with 91% (primary) and 88% (secondary) of that talk being in English. The perceptions studies had shown positive impacts on students' and teachers' confidence and motivation.

All the reports on the research studies referred to in the previous section include a detailed exposition of their respective sampling strategies, with sampling in each case designed to ensure statistical comparability. Since the populations changed dramatically between the studies of Phase II (c.700 teachers) and the first cohort of Phase III (c.4000 teachers), samples for the latter studies were established on the basis of power analyses. Sampling was been based on stratified random samples of clusters and Upazilas across the 7 administrative divisions of the country.

1.3   Purposes of this Study

There were two purposes for study:

1. To provide evidence of results of EIA in the institutionalised phase of the project (Phase IV) as required by the logframe and the AR 2014 comment on the need for a quasi-experimental approach.[18]
2. To provide international evidence for the efficacy of the EIA model of teacher professional development (see *Study Design,* below).

Providing evidence of results from EIA was the primary purpose, with the contribution to the wider international evidence base being an additional benefit.

---

[18] This recommendation 'In general, a limitation of EiA is the lack of a counterfactual or a quasi-experimental design.' (Annual Review , 2014, p. 4)

Evidencing Project Logframe Requirements

There were two levels of the logframe that were relevant to this study of the outcomes of EIA.

Outcome Level

OM1a: Number of people with improved communicative English Language Competence (students)[19]

The milestones and targets of OM1a were expressed in terms of the English Language Competence (ELC) of primary and secondary students measured in terms of the Trinity College London Graded Examinations in Spoken English (GESE):

Primary: ≥ GESE 1: +5%
                ≥ GESE 2: +5%
Secondary: ≥ GESE 2: +10%
                ≥ GESE 3: +5%
Where the percentages were expressed above the baseline established in 2009/2010.

Output Level:

O1b & O3b[20] – Classroom Practice Implementation: The numbers of teachers using Communicative Language Teaching (CLT) approaches in their classroom practice, monitored by % of student talk in lessons and % of that student talk that is in English.

As the output statement above implies, this requires observational data of the amount and language of talk of students in the classroom, and the milestones and targets are:

Student talk: 20% (of the lesson)
Student talk In English: 60% (of student talk)
(These figures apply to both primary and secondary classrooms)

Contributing to international evidence

As noted above, the theory of change of EIA was based upon good evidence from the literature (Westbrook et al., 2013) on the elements that make up its professional development package (the five pillars), but the evidence base was nevertheless relatively limited as the recent DFID review makes clear:

> … students' learning outcomes as a result of ITE (Initial Teacher Education) or CPD (Continuing Professional Development) were often not obtained for reasons of scale and feasibility. Thus, evidence on the impact of training was only partially captured in most studies, highlighting a need for more holistic and robust evaluations of teacher education initiatives (Westbrook et al., 2013, p. 31)

Thus, although the prime purpose was to provide results evidence for the success of EIA, an additional purpose was to contribute to the international evidence base through a holistic and robust evaluation of CPD.

---

[19] In earlier phases of EIA there was an adult learning component and thus the use of the generic 'people' in the outcome statement, but now only the schools component (primary and secondary).
[20] Output 1 refers to primary schools and Output 3 to secondary schools

2.   Methodology

2.1   Study Design

School as a focus of activity

As indicated in the earlier sections, the model of professional development of teachers focuses on the school, through teacher activity in the classroom, and peer and head teacher support (along with further peer support at cluster meetings). Thus, the school is the unit of focus for research. Any study of the improvements that EIA produces in student learning outcomes should relate to the particular school, teacher and classroom. Although in the initial planning stage it was hoped to link classroom data (e.g. teacher ELC and classroom practice) to the students' English Language competency (ELC), in the event the sample sizes necessary to do this, were such that the study would have been too expensive.

There was also a case for collecting additional data to provide explanatory power. For the main sample, this was limited to basic demographic data (location: rural/urban/per-urban; gender). As recommended by SARH, a very small-scale qualitative study was designed to examine the nature and experience of support from peer teachers and head teachers and is reported separately.

Experimental design

Given the Annual Review recommendation to conduct a quasi-experimental study and, following advice from SARH and SEQAS, this study adopted a 'between-groups' approach combined with a 'difference-in-differences' indicators from pre- and post-test studies of control and treatment groups.

Control and treatment groups

Following the description of the National Teacher Development Programmes earlier and in-line with SEQAS recommendations, the control and treatment groups are set out in
Table 1: Composition of the control and treatment groups in terms of existing Teacher Development Programmes.

Table 1: Composition of the control and treatment groups in terms of existing Teacher Development Programmes

	Control	**Treatment (Participating in EIA SBTD intervention)**
Primary schools	2 Teachers of English, 1 of whom took part in SBT in 2014[21]	2 Teachers of English, 1 of whom took part in SBT in 2014
Secondary schools	2 teachers of English	2 Teachers of English

---

[21] National 'Subject-Based Training (SBT)' planned to reach one teacher from each government primary school. Choosing teachers (and therefore schools) who had already completed this training prior to the study was intended to prevent contamination.

Variables and Instruments

There are two groups of variables:

1. Those contained in the two prime instruments:
   a. Students' communicative English Language Competence (ELC), assessed through Trinity College London's Graded Examinations in Spoken English (GESE)
   b. Classroom practice implementation, assessed through timed observations of teacher and student talk during English language lessons.
2. Demographic data that are routinely collected as part of the data collection for '1' (e.g. rural/urban; division; class grade; teacher gender, age and highest qualification).

1a: Students' English Language Competence

ELC was assessed by Trinity College London assessors who carried out a face-to-face protocol in the form of conversation with an individual taking the student through to the highest level at which he or she was able to perform. From this, subjects are assigned to a grade (1-12) that represented the communicative competence of the individual. In the case of students, assessors visited schools and assessed a sample of the students in the classes of the EIA teachers (20 students).

As will be evident under practical issues below, the baseline to end-line study had been planned to be just less than one year of the Programme operation. Normally this would be considered too short a time for any impact on learning outcomes to be revealed, but experience from previous studies on EIA indicate that this statistically significant changes could be obtained. In the event, the treatment time was much less. Further consideration of the practicalities of the implementation time and the timing of the end-line is given below.

1b: Classroom Practices

The observation schedule used recorded a number of elements of classroom practice (CP) which were required for the logframe (i.e. % of student talk and % of this talk in English), and these are reported. Other variables from the schedule are investigated as additional variables, for example:

- the proportions of student talk in pairs and groups (student-student talk);
- the proportions of student reading, writing and listening to audio;
- the proportions of teacher talk in presenting, organising, asking questions and giving feedback.[22]

2: Demographic data

The following demographic data were collected during the classroom observation: administrative division; class grade; the number of boys and girls attending on the day for each class observed. The administrative division can also be related to the types of geography (government classification as urban, rural or peri-urban) and hence general socio-economic conditions.

Data Analysis

Where possible, data analysis was conducted in terms of Difference-In-Differences indicators and of Effect Size, to enable comparison with the internationally accepted findings on innovations (see Annex 2). Where there was insufficient power to express results in terms of statistical significance these are presented as descriptive statics, and clearly indicated as such.

---

[22] There is also an 'other' category, with a selection of possible behaviours such as 'using the blackboard' and 'checking students' work but currently these details are not recorded.

Ethics

As part of normal ethical procedures adhered to by EIA, prior permission was obtained from the head teachers, the teachers and the students to undertake the research and include them in the sample. This involved a written Project Information Sheet, explained by a local speaker in Bangla, the mother tongue of most participants. For teachers and head teachers, the participant information was given in writing and written consent obtained. Written consent for students was provided by teachers and head teachers, who were legally responsible for students whilst they were in school. All participants were made aware that participation was optional and voluntary. All information within the EIA project is held under strict confidentiality and all teachers and students assessed (and their schools) are anonymous in any reporting.

Supplementary Qualitative Study

During negotiation of study design with DFID's SEQAS, the study design team were advised that whilst the project already had strong evidence (from prior research, monitoring and evaluation data) of what happened in teacher development meetings, in classrooms and in effects on students' learning, one of the most under-researched aspects of the Programme related to teachers' experiences of 'support in school' for ongoing development of more communicative teaching and learning practices:

> *"**The concept of support is not straightforward** (different approaches will work for different people) **and is unsuited to quantitative data collection** systems… Instead **I would advise a qualitative approach**, perhaps by **selecting 2 treatment schools and 2 control schools (at both primary and secondary level)** equally split between **good and poor performers** and conducting an in depth assessment at each **to determine what role peer and head teacher support played… if any**."*

In response to this advice, it was agreed to carry out a small-scale qualitative study along the lines suggested, to supplement this large-scale quantitative work. The Phase IV qualitative study is reported separately.

2.2   Methods

The two main measurements are of ELC through a focus on an individual's listening and speaking competence, reflecting the communicative approach of EIA, and Classroom Practice through the systematic observation of lessons.

English Language Competence

As noted earlier ELC testing was conducted by Trinity College London, using its internationally recognised process, and is independent of the EIA team. Assessments took the form of one-to-one, face-to-face oral interviews, carried out by an independent assessor. The assessment "replicates real-life exchanges in which the candidate and the examiner pass on information, share ideas and opinions and debate topical issues" (Trinity College London 2009: p.6).

The assessment was conducted through an interview, the core of which is a conversation element. This is described as "a meaningful and authentic exchange of information, ideas and opinions, rather than a formal 'question and answer' interview." (Trinity College London 2009: p.7). Discussion topics were selected for their potential to elicit the candidate's highest level of EL competence and offer a progression from the familiar to the less familiar and from the 'concrete' to the 'abstract'.

The assessor sought to elicit and facilitate communicative skills, language functions and language items relating to progressively higher grades, ending the interview when the candidate is judged to have reached the peak of his/her capacity. At this point the candidate was assigned a Trinity grade (1–12).

It is a valid and internationally recognised assessment of both English Language Competence (through its benchmarking to the Common European Framework of Reference (CEFR) for Languages, see Trinity College London 2007) and of the specific communicative approach to English Language teaching (ELT) promoted by EIA (through the use of the 'conversational' approach indicated above). The reliability of the assessment is ensured by the international experience and high levels of staff training and moderation (Trinity College London 2013).

In the past ELC testing by Trinity assessors has been a combination of visits to schools and visits to teacher cluster meetings. In schools they test students and the two EIA teachers. For the QE study, Trinity assessors carried out assessments in one school per day, and hence needed 120 assessor-days to cover the sample of 120 schools.

Classroom Practice observation

The observation schedule used by EIA focused on talk as this is the main focus of the communicative approach (although the EIA materials include all four EL skills: listening, speaking, reading and writing), but with some record of non-talk activity. The schedule involved systematic observation using instantaneous sampling at one-minute intervals for the whole lesson. This was done by trained researchers (see the next section on *Fieldwork*), but the level of judgement was minimised to maximise the reliability of the measure.[23]

EIA has worked with the Institute of Education and Research (IER), Dhaka University for many years, drawing on it for the field workers for observation, questionnaire administration and interviews of teachers and students. The field workers were MPhil students supported by EIA who, as part of their studies, spent time working in the field as researchers. For the QE study, IER provided 10 new MPhil students from January 2015. The IER fieldworkers observed on average in 1 school (2 teachers, 1 lesson per teacher) per day, requiring 120 observer-days or 12 days for each observer. This meant that data collection for the classroom observation took about 2 weeks.

Reliability measure for Classroom Practice

Agreement of the observers was assessed via Krippendorffs' $\propto$ (alpha) (Krippendorff, 2004). An acceptable level of agreement was determined to be as $\propto \geq 0.8$, with results being treated as tentative if $\propto$ (alpha) is in the range 0.667-0.8.

Inter coder reliability was calculated by using Krippendorff's macro in SPSS, using an updated version of the macro provided directly by the macro-author to the data analyst (this macro fixes a bug found in the version in the public domain). The resulting Krippendorff's α = .7849, fell just below but rounded to the pre-determined acceptable level of agreement.

---

[23] It is based on an instrument used by other ELT researchers.

2.3 Sample

EIA used a stratified random sampling (applied independently for primary and secondary Upazilas and schools, to avoid bias) with selected schools being randomly allocated to control or experimental groups. This approach enabled a 'between groups' approach, using difference-in-differences indicators, in accordance with SEQAS advice. In total the same 120 schools (and hence 240 teachers and 2,400 students[24]) were involved in the studies, as shown in Table 2: Planned samples of intervention and control schools, teachers and students.[25]

Table 2: Planned samples of intervention and control schools, teachers and students

	Pre-intervention			Post-intervention		
Primary	Schools	Teachers	Students	Schools	Teachers	Students
Intervention group	30	60	900	30	60	600+
Control group	30	60	900	30	60	600+
Secondary						
Intervention group	30	60	900	30	60	600+
Control group	30	60	900	30	60	600+
Total	**120**	**240**	**3,600**	**120**	**240**	**2,400+**

A stratified random sampling was used to identify a number of Upazilas per division, with one control and one experimental school being selected from each Upazila. Sampling, from division, to school, to Upazila, was carried out independently, for primary and secondary phases (Table 3 and Table 4).

Table 3: Sampled primary schools per division (total EIA upazilas-210)

Division	Number of EIA intervention Upazilas per division	Percentage of EIA Upazilas per division	Target number of 60 sample schools per division	Even number of schools (control & treatment) per division	Number of Upazilas to sample
Chittagong	34	16.2	9.7	10	5
Rajshahi	26	12.4	7.4	8	4
Khulna	30	14.3	8.6	8	4
Sylhet	18	8.6	5.1	6	3
Barishal	17	8.0	4.9	4	2
Dhaka	59	28.1	16.8	16	8
Rangpur	26	12.4	7.4	8	4

---

[24] Actually 3,600 students (15 per teacher) were assessed at pre-test, to maximize the chances of getting 2,400 of the same students (10 per teacher) present for both pre- and post- test assessments.
[25] Actual samples are given in the Findings section 7.1.

In primary, the EIA intervention took place in 210 Upazilas, with the Upazila list being negotiated and agreed with the GoB. The number of Upazilas from each division to be included in the QE sample, was chosen to be proportionate to the number of EIA Upazilas in that division. Following random sample from division to Upazila, the following primary Upazilas were sampled:

Chittagong: Hathazar, Chandgoa, Brahmanpara, Chauddagram, Companigonj (5)
Rajshahi: Lalpur, Sonatola, Dhunat, Raiganj (4)
Khulna: Gangni, Khulna Sadar, Alamdanga, Shymnagar (4)
Sylhet: Tahirpur, Habiganj Sadar, Bahubal (3)
Barishal: Barguna Sadar, Patuakhali Sadar (2)
Dhaka: Mirpur, Keraniganj, Dhamrai, Raipura, Ghior, Shibchar, Mirzapur, Nagarpur (8)
Rangpur: Dimla, Boda, Palashbari, Patgram (4)

Table 4: Sampled secondary schools per division (total EIA upazilas-100)

Division	Number of EIA intervention Upazilas per division	Percentage of EIA Upazilas per division	Target number of 60 sample schools per division	Even number of schools (control & treatment) per division	Number of Upazilas to sample
Chittagong	16	16	9.6	10	5
Rajshahi	12	12	7.2	8	4
Khulna	14	14	8.4	8	4
Sylhet	9	9	5.4	4	2
Barishal	9	9	5.4	6	3
Dhaka	28	28	16.8	16	8
Rangpur	12	12	7.2	8	4

Following random sample from division to Upazila, the following secondary Upazilas were sampled:

Chittagong: Ramgrh, Kaptai, Raozan, Cox's Bazar Sadar, Faridganj (5)
Rajshahi: Natore Sadar, Raiganj, Chapai Nawabganj Sadar, Rajshahi Sadar (4)
Khulna: Magura Sadar, Narail Sadar, Keshabpur, Shyamnagar (4)
Sylhet: Fenugonj, Habiganj sadar (2)
Barishal: Barisal Sadar, Jhalokati Sadar, Pirojpur Sadar (3)
Dhaka: Ghior, Narsingdi Sadar, Munshiganj Sadar, Rajbari Sadar, Madaripur Sadar, Kotalipara, Gaffargoan, Mirzapur (8)
Rangpur: Gangachara, Pirgonj, Sundargonj, Dinajpur Sadar (4)

From each of the Upazilas above, two schools were randomly sampled. One school was randomly allocated to the control group, the other to the treatment group. From each school (control and treatment alike) two teachers of English language were chosen. In primary, the English teacher from each school who received SBT in 2014 was selected, along with one other teacher of English. In secondary, the two teachers who teach the most English were selected. For each teacher, one class was selected for lesson observation; from this class, 15 students[26] were chosen at random for English Language Competence (ELC) testing. An even distribution of class grades were chosen. The post-test observed the same classes and assess the same students, as the pre-test, as far as possible.[27]

---

[26] These numbers made some allowance for attrition (there is no *statistical power* reason to increase the number of students above 10); the intention was to increase the likelihood of achieving 10 particular individual students per teacher, who are present for both pre- and post-tests.

[27] Where there were no pre- and post-test data, then these cases were omitted from the analysis (Tables 7-12 give the full experimental sample data).

2.5 Timeline

Timeline by Design

Figure 1: Timeline as per final study design

Strand			May 15	Jun 15	Jul 15	Aug 15	Sep 15	Oct 15	Nov 15	Dec 15	Jan 16	Feb 16	Mar 16	Apr 16	May 16	Jun 16	Jul 16
Bangladesh School closures		Primary			Ramadan			Revision: limited teaching		Schools Closed	Admin: limited teaching						Ramadan
		Secondary															
EIA Phase IV SBTD Programme	Cohort 1	Primary	Launch (CM1)			CM2					CM3			CM4			
		Secondary										CM4					
	Cohort 2	Primary												CM1			CM1
		Secondary											CM 1				
EIA QE Research Study			Pre-test											Post-test		
Supplementary Qualitative Study								Fieldwork				Fieldwork				

Forthcoming year May 2015 – July 2016

The intention in initial design, was for the study to take place in one school year, with all School-Based Teacher Development (SBTD) activities taking place between March – September 2015, with pre- and post- test just before and after these dates. Nationwide security disruption in the first months of 2015 forced the planned SBTD Programme to be pushed back, straddling two school years, 2015 and 2016. This was far from ideal in terms of the study, as some teachers may be posted to different schools at the start of the school year and in larger schools, it is possible for students to end up in a different section (teaching group) or to be taught by different subject teachers. However, at the time plans had to be finalised, it was not possible to postpone the study to the following year, as the funding for academic contribution to EIA was due to end in July 2016. This situation was reviewed with South Asia Research Hub and DFID Bangladesh and the study design amended for pre-test in May 2015 and post-test in May 2016, with the SBTD Programme intended to complete in these dates.

13

Timeline Achieved

Figure 2: Timeline (actual achieved)

Strand			May 15	Jun 15	Jul 15	Aug 15	Sep 15	Oct 15	Nov 15	Dec 15	Jan 16	Feb 16	Mar 16	Apr 16	May 16	Jun 16	Jul 16	Aug 16	Sept 16	Oct 16
Planned and Actual Timelines: Cohort 1 Implementation Period: May 2015 – October 2016																				
Bangladesh School closures	Primary				Ramadan			Revision: limited teaching		Schools Closed	Admin: limited teaching									
	Secondary																			
EIA Phase IV SBTD Programme	Cohort 1 Plan	Primary	Launch (CM1)			CM2					CM3			CM4						
		Secondary											CM4							
	Cohort 1 Actual	Primary				Launch CM 1		Launch CM 1 Continued			CM 2			CM 3			CM 4			
		Secondary																		
EIA QE Research Study			Pre-test												Post-test					
Supplementary Qualitative Study								Fieldwork											Fieldwork

In practice, the launch of the SBTD Programme was further delayed by nationwide demonstrations and protests by the opposition party, lasting several months. Once security was normalised a number of activities (such as preparatory workshops with head teachers, teacher facilitators and education officers) had to be completed before launch of Cluster Meeting 1 (CM1) in August-November. Once started, the operational plan was for CMs to be led by Teacher Facilitators; in practice, divisional teams led almost all CMs, causing delays to completion of each CM cycle.

For five months after launch (mid-September to February), teachers had little or no opportunity to apply training to practice, due to holidays, revision, national examinations and school closures. In effect, CM 2 became a 'relaunch', occurring six-months, rather than six weeks, after CM 1. With the 'relaunch' of CM2 between January and March, the only effective 'teaching time' available for the introduction of new classroom activities was limited to around two-to-three months, from February/March to April/May 2016. Teachers were at best, effectively only half way through the SBTD treatment at post-test. It was not possible to put the study back beyond May 2016, as at the time contracts for fieldwork and logistics had to be agreed, funding for academic input to EIA (including the reporting of this study) was due to finish in July 2016, with full project closure by March 2017.

3   Findings

These findings are presented in four parts in terms of:

1. the samples of the ELC assessment of students and the classroom practice observations,
2. the logframe requirements of outcomes (increased proficiency of communicative English) and outputs (teachers' use of CLT in their classrooms);
3. further descriptions of classroom practices;
4. demographic effect (gender and location)

3.1   Achieved Samples

Student assessment: ELC competency

The actual samples of students who were tested at both pre- and post-treatment ELC assessments were 75% of that planned for students (1,802 of 2,600 planned). 72% of the planned for teachers' lesson were observations (163/240), with the difference attributed to security issues that caused the study to span more than one school year (see Table 5 and explanation), with associated churn in the teacher and student populations. The detailed samples of students assessed and classrooms observed are given in the Findings (*see 3.1   Achieved Samples*). The samples were sufficient for the planned analyses across the experimental conditions.

This sample has been achieved despite substantial political and security instability across the country causing delays to the start of the treatment and disruption to the pre-test fieldwork. These delays also necessitated the duration of the experimental study running over the transition from one school (and calendar) year to another, in order to allow time for the treatment to take place. Running across two school-years led to higher than anticipated 'churn' in both the teacher workforce deployment and the student population. That the study still managed to achieve 75% of the planned student sample and 72% of the planned teacher sample present for both pre- and post-test, is a result of very extensive and pro-active efforts to maximise availability of participants at the post-test.

Of the excluded students, around 89% of these were because the students were unavailable (absent or no longer in the class/school) at the time of post-test fieldwork.

Most of the remaining 11% or so were excluded due to a rigorous sample ratification process prior to post-test:

- 8 Schools assigned as 'control' were included in the treatment because of miscommunication at field level. These were excluded from the study;

- 10 teachers and their students were excluded based on pre-fieldwork telephone screening, indicating the teachers could no longer participate (e.g. teachers were transferred, on maternity leave, on long-term sick leave, or assigned to non-teaching duties).

Table 5 gives the samples of those students who were assessed for ELC (GESE) at both pre- and post-test, across both control and treatment groups. Table 6 breaks these data down into the two groups and by phase (primary and secondary). Table 7 gives the breakdown of the reasons for exclusion of students. This table shows that 89% of students who were excluded from the study were those who

were absent at the post-test fieldwork. Their scores were either left blank (37%) or marked absent (A) (52%), dependent upon the practice of the fieldworker. A further 11% of students were excluded as a result of the rigorous verification process following pre-test.

Table 5: Actual sample for student ELC assessments (GESE)

Sample	N (%)
Pre test	3484 (100%)
Post test	1802 (51.7%)
Excluded	1682 (48.3%)

Table 6: Actual sample for student ELC assessments (control and treatment by phase)

	Primary	Secondary	Total
Treatment (EiA)	620	482	1102
Control (Non-EiA)	369	331	700
Total	989	813	1802

Table 7: GESE Excluded data

Exclusion rationale	Frequency	Percentage
Post-test is absent (A)	1499	89.1
Withdrawn from study*	180	10.7
Outlier	2	0.1
No pre test data	1	0.1
Total	**1682**	**100.0**

*180 students from 14 schools were withdrawn after rigorous review of initial sample;

Classroom Practice Observations

As with ELC: Table 8 gives the sample sizes for all classrooms observed at both pre- and post-test, across both control and treatment groups; Table 9 gives the breakdown of classrooms (teachers) observed for both groups and across phases (primary and secondary); and Table 10 indicates the reasons for exclusion of data in the same way as done in Table 7.

Table 8: Actual sample for Classroom Practice observations (classrooms)

Sample	N (%)
Pre test	242 (100%)
Post test	163 (67.4%)
Excluded	79 (32.6%)

Table 9: Actual sample for classroom practice observations (control and treatment by phase)

	Primary	Secondary	Total
Treatment (EiA)	45	49	94
Control (Non-EiA)	34	35	69
Total	79	84	163

Table 10: Classroom practice observations, excluded data (classrooms)

Exclusion rationale	Frequency	Percentage
No post-test data	72	91.2
Withdrawn from study*	5	6.3
No pre-test data	2	2.5
Total	**79**	**100.0**

* five teachers were withdrawn from the study, for example, if they were on sick-leave or no-longer teaching English.

3.2   Logframe Requirements: Post-test results, Treatment Schools Only

This part of the results examines the logframe requirements of both ELC assessment and classroom practice (CP). The ELC assessment is an Outcome indictor and CP is an output indicator.

Output Indicators O1b & 3b[28]

Implementation: The numbers of teachers using Communicative Language Teaching (CLT) approaches in their classroom.

These indicators are expressed specifically in terms of 'student talk' as proportion of the lesson time and proportion of that 'student talk in English' and, for each, the logframe defines targets. Table 11 gives the achievements of the EIA intervention schools in both primary and secondary, and Table 12 compares these with the targets. As Table 12 indicates, the targets, bar one, are exceeded and in some cases by a considerable amount.

Table 11: Logframe Communicative Language Teaching (CLT) results

Phase	CLT measure	Percentage
Primary (N=45)	Student talk/lesson-time	24.6
	Student talk in English/student talk-time	91.3
Secondary (N=49)	Student talk/lesson-time	19.7
	Student talk in English/student talk-time	84.5

---

[28] O1 (Output 1) refers to primary and 03 (Output 3) refers to secondary.

Table 12 Logframe CLT results against targets

School		Percentage	Target	Difference
Primary (N=45)	Student talk/lesson-time	25	20%	+5%
	Student talk in English/student talk-time	91	60%	+31%
Secondary (N=49)	Student talk/lesson-time	20	20%	0
	Student talk in English/student talk-time	85	60%	+25%

Table 12 shows that in all cases, EIA has achieved or exceeded the output indicator targets (O1b and 03b), with very substantial improvements over target for half of the indicators.

Outcome Indicator OM1a:

Increases in proficiency in communicative English among the population of Bangladesh (students)

The logframe sets two criteria for student competency in terms of the GESE grades (Trinity) in each of the two interventions by EIA (primary and secondary). The intention is that there should be an improvement in the proportion of students achieving at or above the criteria grades. In Table 13: Logframe Results, Improvements over pre-test in EIA schools (GESE), these criteria are shown (column 1) and the actual scores measured pre- and post-test are given along with associated differences and statistical significance. It is evident that in both phases there are statistically significant improvements in the EIA scores from pre- to post-test, for all criteria.

Table 13: Logframe Results, Improvements over pre-test in EIA schools (GESE)

	Criteria (Grade)	Pre-test	Post-test	Difference	*p*
Primary	≥ GESE 1	39.8%	58.9%	+19.1%	*P < .001*
	≥ GESE 2	9.0%	11.3%	+2.3%	*P < .001*
Secondary	≥ GESE 2	40.0%	51.5%	+11.5%	*P < .001*
	≥ GESE 3	26.3%	28.0%	+1.7%	*P < .001*

The EIA school pre- and post-test GESE scores are shown in bar charts for primary (reference) and secondary (reference) below.

Figure 3: Primary Student GESE Scores, EIA Treatment

Figure 4: Secondary Student GESE Scores, EIA Treatment

The logframe also contains targets for improvements (over the baseline) in each phase to be achieved in 2017 against these two criteria.

Table 14 indicates the actual improvements compared to the targets in the logframe. This shows that at the lower grade criteria, primary achievement is 14% points above target and secondary achievement is 2% points above target; at the higher grade criteria the EIA intervention showed 2% points improvement in primary and in secondary, but both phases were 3% points below the targets. This indicates that, as has been apparent over the years of the cohort studies, EIA does better in improving students at the lower grades. However, given outcome targets were for achievement *at the end of the treatment*, but due to nationwide protests schools were only halfway

through the treatment at post-test, this evidence suggests it is likely that all targets would have been met or exceeded if assessments had been made at the end of the treatment.

Table 14: ELC results (GESE) compared to Logframe targets by phase

	Criteria	Target	Achieved	Difference
Primary	≥ GESE 1	+5%	+19%	+14%
	≥ GESE 2	+5%	+2%	-3%
	Total gains	+10%	+21%	+11%
Secondary	≥ GESE 2	+10%	+12%	+2%
	≥ GESE 3	+5%	+2%	-3%
	Total gains	+15%	+14%	-1%

3.3 Experimental Study Deliverables: Difference-In-Differences (DID) indicators and Effect Sizes

In this section, the results of the comparisons of the EIA treatment group with the control group are given for both ELC and classroom practice in each of the two phases (primary and secondary). These results are examined in terms of the improvements in pre- and post-test for each group (difference) and the comparison of these improvements for the control group and treatment group (difference in differences). These latter figures are examined in terms of statistical difference and expressed as an effect size.[29]

Students' English Language Proficiency (GESE grade)

These data give the pre- and post-assessment of ELC for students for both the control and the EIA treatment groups and, as indicated above, presents the difference in differences of the two groups and its statistical significance and the resulting effect size. Table 15 presents these data for all students assessed and by phases (primary and secondary), with the ELC expressed in terms of the mean score (GESE grade) for each group. All students and those in each phase show improvements for pre- to post-test, but the difference in differences is zero or very small (0.1), and none are statistically significant. Consequently, few show a meaningful effect size (< 0.04). As the pre-/post-test differences are positive Table 16, Table 17 and Table 18 examine the statistical significance of these differences (the difference in differences are also given). The sample sizes (N) for each are given in parenthesis. These tables show that in all cases the differences are statistically highly significant.

---

[29] The effect size is calculated as a Pearson coefficient (*r*), and the explanation is found in Appendix 2.

Table 15: Students Proficiency (GESE), summary experimental results

All/Phase	Control			Treatment			Difference in Differences	Statistical significance (p <0.05) (1-sided)	Effect Size r
	Pre	Post	Difference	Pre	Post	Diff			
All students	1.1	1.3	0.2	1.0	1.3	0.3	0.1	$p = .095^{ns}$	.03
Primary students	0.5	0.7	0.2	0.5	0.7	0.2	0.0	$p = .270^{ns}$	.02
Secondary students	1.8	2.0	0.2	1.7	2.0	0.3	0.1	$p = .116^{ns}$	.04

Table 16: ELC (GESE) All students (N=1802)

	Pre-test	Post-test	Difference
EiA (N=1102)	1.0	1.3	$0.3^{***<.001}$
Non-EiA (N=700)	1.1	1.3	$0.2^{***<.001}$
Difference	-0.1^{ns}	0.0^{ns}	**DD = 0.1^{ns}**

2-sided $p = .189$, 1-sided $p = .095$, $r = .03$, Note: significant differences <u>in</u> the table are tested one-sided.

Table 17 Primary students GESE (N= 989)

	Pre-test	Post-test	Difference
EiA (N=620)	0.5	0.7	$0.2^{***<.001}$
Non-EiA (N=369)	0.5	0.7	$0.2^{***<.001}$
Difference	0.0^{ns}	0.0^{ns}	**DD = 0.0^{ns}**

2-sided $p = .539$, 1-sided $p = .270$, $r = .02$, Note: significant differences <u>in</u> the table are tested one-sided.

Table 18: Secondary students GESE (N= 813)

	Pre-test	Post-test	Difference
EiA (N=482)	1,7	2,0	$0,3^{***<.001}$
Non-EiA (N=331)	1,8	2,0	$0,2^{**.002}$
Difference	$-0,1^{ns}$	$0,0^{ns}$	**DD = 0,1^{ns}**

2-sided $p = .231$, 1-sided $p = .116$, $r = .04$, Note: significant differences <u>in</u> the table are tested one-sided.

As noted above, there are improvements in students' proficiency across the board, under all conditions. Though modest in terms of whole GESE grades, these improvements are highly statistically significant (because they are relatively large compared to both the mean pre-test levels and the variation within the population and they are seen across a large sample size).

However, there was little evidence of an experimental effect (expressed as DID or effect size) as there were no statistically significant differences between the improvements of the control and treatment groups.

Classroom practice

Here student talk time is examined, along with the percentage of that time where talk was in English.

Student talk time (as percentage of lesson time)

These data are from the classroom observations and express the schedule variable 'student talk' in terms of a mean percentage of the overall lesson time. Table 19 presents these data by phases (primary and secondary), with the student talk expressed as a percentage of the lesson time for each group. This table indicates that there were improvements in both phases of the treatment groups, but not in both of the control groups (primary had a reduced 'difference'). However, again there were no instances of statistically significant difference-in-differences for either phase and the resulting effect sizes were nominal to zero.

Table 19: Student Talk time as a percentage of lesson time, summary experimental results

	Control			Treatment			Diff in Diffs	Statistical significance ($p < 0.05$) (1-sided)	Effect Size r
Variable	**Pre**	**Post**	**Difference**	**Pre**	**Post**	**Diff**			
Primary %	24.3	23.6	-0.7	22.8	24.6	1.8	2.5	.248[ns]	.08
Secondary %	15.9	18.0	2.1	17.5	19.7	2.2	0.1	.495[ns]	.00

Again more details of the differences are given in Table 20 and Table 21, where the statistical significance of differences are examined, reinforcing the changes noted above that in all conditions, student talk-time was relatively high at pre-test and in most conditions increased slightly at post-test, and in addition the tables indicate that none of the differences are statistically significant in either phase. Given the high student talk at pre-test, the educational significance of any such small changes is likely to be negligible.

Table 20: Percentage student talk time, Primary (N= 79)

	Pre-test	Post-test	Difference
EiA (N=45)	22.8	24.6	1.8[ns]
Non-EiA (N=34)	24.3	23.6	-0.7[ns]
Difference	-1.5[ns]	1.0[ns]	**DD= 2.5[ns]**

2-sided p = .495, 1-sided p = .248, r = .08

Table 21: Percentage student talk time, Secondary (N= 84)

	Pre-test	Post-test	Difference
EiA (N=49)	17.5	19.7	2.2^{ns}
Non-EiA (N=35)	15.9	18.0	2.1^{ns}
Difference	1.6^{ns}	1.7^{ns}	**DD= 0.1^{ns}**

2-sided p = .989, 1-sided p = .495, r =.00

Students' talk in English (as percentage of student talk-time)

These data take the student 'talk time' from the previous section and present the data on the proportion of this talk that is in English (expressed as a percentage). Table 22 presents these data by phases (primary and secondary), with the student talk in English expressed as a percentage of the total student talk time for each of treatment and control group (note that 'all classrooms' are not included, as the classroom practice is necessarily different in the two phases).

Table 22: Student talk in English as percentage of student talk-time, summary experimental result

	Control			Treatment			Difference in Differences	Statistical significance (p <0.05) (1-sided)	Effect Size r
Variable	*Pre*	*Post*	*Difference*	*Pre*	*Post*	*Diff*			
Primary Percentage	81.1	91.2	10.1	87.9	91.3	3.4	-6.7	p = .139 ns	.12
Secondary Percentage	73.9	66.2	-7.7	77.0	84.5	7.5	15.2	p = .040 *	.19

Note, data marked '*' indicates statistically significant.

In primary, there is a negative experimental effect in DID in Table 22 (e.g. the control group improved more than treatment), however, both conditions ended up at the same percentage of student talk in English, whilst the control group had a lower pre-test value. The difference in differences was, however, not statistically significant, and essentially this means that nothing can be concluded apart from the fact that there was no experimental effect.

In secondary, there is a large and statistically significant DID in secondary student percentage talk in English (Table 25). This is the first experimental effect, and it has a small-to-medium effect size (Pearson r=0.2).

Table 23: Student's talk in English, all classrooms (N=163)

	Pre-test	Post-test	Difference
EiA (N=94)	82.2	87.8	5.6*[.033]
Non-EiA (N=69)	77.4	78.5	1.1^{ns}
Difference	4.8^{ns}	9.3*[.020]	**DD= 4.5^{ns}**

2-sided p = .403, 1-sided p = .202, r = .07
Note: significant differences in the table are tested one-sided.

Across primary and secondary as a whole, students talk in English increases in EIA schools and this is statistically significant, whereas there is a much smaller increase in control schools, which is not statistically significant. Whilst there is a positive difference in differences, it is not statistically significant. Table 24 and Table 25 give more detail on the statistical significance of the differences in the pre- and post-tests for all classrooms for both primary and secondary (respectively). Table 24 indicates that in primary classrooms there is a negative difference in differences, although it is not statistically significant. However, the improvement in the control group *is* statistically significant. Post-test both groups a similar percentage of English in student talk, but the control group has a lower pre-test percentage of English in student talk than the treatment.

Table 24: Students talk in English, Primary (N= 79)

	Pre-test	Post-test	Difference
EiA (N=45)	87.9	91.3	3.4^{ns}
Non-EiA (N=34)	81.1	91.2	$10.1^{*.014}$
Difference	6.8^{ns}	0.1^{ns}	**DD= -6.7**

2-sided p = .277, 1-sided p = .139, r = .12
Note, significant differences <u>in</u> the table are tested one-sided.

For secondary classrooms (Table 25), there is a statistically significant increase in secondary students' use of English between pre- and post-test, but no-statistically significant difference in control schools. The difference in differences is also statistically significant, with a small-to-medium effect size. This is the statistically significant experimental effect shown in Table 22.

Table 25: Students talk in English, Secondary (N= 84)

	Pre	Post	Difference
EiA (N=49)	77.0	84.5	$7.5^{*.043}$
Non-EiA (N=35)	73.9	66.2	-7.7^{ns}
Difference	3.1^{ns}	$18.3^{**.008}$	**DD= 15.2***

2-sided p = .080, 1-sided p = .040, r = .19
Note, significant differences <u>in</u> the table are tested one-sided.

3.4 Further descriptions of classroom practices

In this section, several other observed variables are examined that are not isolated in the logframe, namely: student-student talk time; teacher talk time, teacher talk time in English, and teacher talk time presenting and organising student activity. Table 26 shows data on these variables.

Table 26: further CP variables, all classrooms (N=163, Primary and Secondary)

Variable	Control			Treatment			Difference in Differences	Statistical significance (p <0.05) (1-sided)	Effect Size r
	Pre	**Post**	**Diff**	**Pre**	**Post**	**Diff**			
Student-student talk (% talk time)	8.4	8.7	0.3	10.5	14.4	3.9	3.6	p = .181 ns	.07
Primary	12,0	9,7	-2,3	4,2	15,9	11,7	14,0	p = .006 **	.27
Secondary	5,0	7,8	2,8	16,3	13,0	-3,3	-6,1	p = .122 ns	.13
Teachers' talk time (% lesson time)	53,6	52,2	-1,4	53,0	49,1	-3,9	-2,5	p = .206 ns	.06
Primary	44,7	44,7	0,0	49,8	44,2	-5,6	-5,6	p = .088 ns	.15
Secondary	62,2	59,5	-2,7	55,9	53,6	-2,3	0,4	p = .464 ns	.01
Teachers' talk in English (% talk time)	63,1	74,1	11,0	69,0	86,6	17,6	6,6	p = .099 ns	.09
Primary	64,0	79,6	15,6	75,0	87,1	12,1	-3,5	p = .303 ns	.05
Secondary	62,1	68,7	6,6	63,5	86,1	22,6	16,0	p = .018*	.21
Teachers' talk presenting (%talk time)	46,5	49,6	3,1	48,9	42,6	-6,3	-9,4	p = .012*	.18
Primary	39,3	45,5	6,2	42,8	39,1	-3,7	-9,9	p = .05*	.19
Secondary	53,4	53,6	0,2	54,5	45,9	-8,6	-8,8	p = .062 ns	.17
Teachers' talk organising student activity (%talk time)	19,2	22,4	3,2	20,1	27,0	6,9	3,7	p = .141 ns	.08
Primary	24,0	25,5	1,5	26,1	32,7	6,6	5,1	p = .155 ns	.11
Secondary	14,5	19,4	4,9	14,6	21,7	7,1	2,2	p = .316 ns	.05

Student-to-student talk as a percentage of student talk time, shows that overall there is a difference-in-differences, but that it is not statistically significant. But for primary classrooms, there is a highly statistically significant difference in differences. This is our second experimental effect, with medium effect size (Pearson r >0.2). Secondary classroom show a non-significant negative difference in differences and hence it is safe to assume there was no difference-in-differences.

Teachers talk in English as a percentage of their talk time also shows a statistically significant difference-in-differences for secondary classrooms. This the third statistically significant experimental effect, with small-to-medium effect size (r in the order of 0.2).

Teachers time presenting shows a statistically significant difference-in-differences. This occurs for primary teachers where this time is reduced (something advocated by EIA). This is the fourth statistically significant experimental effect, with a small-to-medium effect size (r in the order of 0.2). Whilst the effect size is similar for secondary teachers, it is just above the threshold for statistical significance. For teacher time organising there is a positive experimental effect for EIA, with small effect size, but this is not statistically significant.

Student-student talk as percentage of student talk-time: differences

In this sub-section, the first variable in Table 26 is examined in more detail to consider the statistical significance of the differences. Student-student talk is made up of pair and group work talk, and across all classrooms there is a modest difference in differences, but it is not statistically significant (Table 27).

Table 27: student-student talk as percentage of all student talk, all classrooms (N=163)

	Pre-test	Post-test	Difference
EiA (N=94)	10.5	14.4	$3.9^{.080}$
Non-EiA (N=69)	8.4	8.7	0.3^{ns}
Difference	2.1^{ns}	$5.7*^{.040}$	**DD= 3.6^{ns}**

2-sided p = .361, 1-sided p = .181, r = .07. Note, significant differences <u>in</u> the table are tested one-sided.

Looking at this interaction by phase shows that in fact the situation in the primary classroom is more positive, with the EIA treatment group making a statistically significant improvement from pre- to post-test (Table 28) and the difference in differences being large and statistically significant. The situation in secondary, however, shows no statistically significant improvement nor difference-in-differences (Table 29).

Table 28: Student-Student talk as percentage of all student talk, Primary (N=79)

	Pre-test	Post-test	Difference
EiA (N=45)	4.2	15.9	$11.7***^{<.001}$
Non-EiA (N=34)	12.0	9.7	-2.3^{ns}
Difference	$-7.8*^{.022}$	6.2^{ns}	**DD= 14.0****

2-sided p = .012, 1-sided p = .006, r = .27 Note: significant differences <u>in</u> the table are tested one-sided, and significance indicated by '*'.

Table 29: Student-Student talk as percentage of all student talk, Secondary (N=84)

	Pre-test	Post-test	Difference
EiA (N=49)	16.3	13.0	-3.3^{ns}
Non-EiA (N=35)	5.0	7.8	$2,8^{ns}$
Difference	$11.3**^{.009}$	5.2^{ns}	**DD= -6.1^{ns}**

2-sided p = .243, 1-sided p = .122, r = .13 Note: significant differences <u>in</u> the table are tested one-sided.

Teacher talk time (as percentage of lesson time)

The EIA approach encourages teachers to increase the amount of student talk, and hence by implication to reduce the amount of their own 'teacher talk' time. This is examined in Table 30, which shows that for the treatment group there was a statistically significant reduction in teacher talk time as a proportion of the whole lesson (with no change in control group), however the difference in differences is not statistically significant.

Table 30: Teachers talk as percentage of lesson time, all classrooms (N=163)

	Pre-test	Post-test	Difference
EiA (N=94)	53.0	49.1	$-3.9*^{.031}$
Non-EiA (N=69)	53.6	52.2	-1.4^{ns}
Difference	-0.6^{ns}	-3.1^{ns}	**DD= -2.5^{ns}**

2-sided p = .412, 1-sided p = .206, r = .06 Note: significant differences <u>in</u> table are tested one-sided.

When the two phases are examined separately, the primary classrooms again show a statistically significant reduction in the teacher talk time (Table 31), and is close to showing a statistically significant difference-in-differences, with low effect size of 0.15 (the control group shows no change from pre- to post-test). Again, the secondary phase shows no improvement and no difference-in-differences (Table 32).

Table 31: Teachers Talk, Primary (N=79)

	Pre-test	Post-test	Difference
EiA (N=45)	49.8	44.2	$-5.6*^{.031}$
Non-EiA (N=34)	44.7	44.7	0.0^{ns}
Difference	$5.1^{.055}$	-0.5^{ns}	**DD= -5.6^{ns}**

2-sided p = .175, 1-sided p = .088, r = .15 Note: significant differences <u>in</u> the table are tested one-sided.

Table 32: Teachers' Talk, Secondary (N=84)

	Pre-test	**Post-test**	**Difference**
EiA (N=49)	55.9	53.6	-2.3[ns]
Non-EiA (N=35)	62.2	59.5	-2.7[ns]
Difference	-6.3*[.037]	-5.9*[.046]	**DD= 0.4[ns]**

2-sided p = .928, 1-sided p = .464, r = .01 Note: significant differences <u>in</u> the table are tested one-sided.

Teachers' talk in English (as percentage of teacher talk-time)

Whatever the level of teacher talk, EIA encourages them to use as much English as possible, while ensuring that students understand both the activities they do and any explanations given. Table 33 presents the data for all classrooms. This shows that both the control and treatment groups made a statistically significant improved in the amount of English spoken by the teacher, but that this improvement was greater for the treatment group. However, the difference in differences was not statistically significantly large.

Table 33 Teachers talk in English, all classrooms (N=163)

	Pre-test	**Post-test**	**Difference**
EiA (N=94)	69.0	86.6	17.6***[<.001]
Non-EiA (N=69)	63.1	74.1	11.0**[.007]
Difference	5.9[ns]	12.5**[.002]	**DD= 6.6[ns]**

2-sided p = .198, 1-sided p = .099, r = .09 Note: significant differences <u>in</u> the table are tested one-sided.

When this variable is examined for each of the phases it is evident that, for primary classrooms (Table 34), the gains from pre- to post-test for the treatment and control groups are reversed from that in all classrooms, i.e. the control group has a larger improvement in the amount of English spoken by the teacher, and that both groups did have a statistically significant increase. It should be noted, however, that the control group (non-EiA) started from a lower base. The difference in differences, although negative for EIA, was not statistically significant.

Table 34 Teachers Talk in English, Primary (N=79)

	Pre-test	**Post-test**	**Difference**
EiA (N=45)	75.0	87.1	12.1**[.002]
Non-EiA (N=34)	64.0	79.6	15.6**[.006]
Difference	11.0*[.038]	7.5[ns]	**DD= -3.5[ns]**

2-sided p = .606, 1-sided p = .303, r = .05 Note: significant differences <u>in</u> the table are tested one-sided.

In contrast, Table 35 shows that only the secondary treatment group showed an improvement (and it was statistically significant) and thus the difference in differences was relatively large, with a small-to-medium effect size of r=0.21.

Table 35 Teachers Talk in English, Secondary (N=84)

	Pre-test	Post-test	Difference
EiA (N=49)	63.5	86.1	22.6***$^{<.001}$
Non-EiA (N=35)	62.1	68.7	6.6^{ns}
Difference	1.4^{ns}	17.4**$^{.005}$	**DD= 16.0***

2-sided p = .035, 1-sided p = .018, r = .21 Note: significant differences <u>in</u> the table are tested one-sided.

Teachers' talk presenting (as percentage of teacher talk-time)

The picture in Bangladesh school when the baseline studies were done for EIA in 2010 was of a lot of the teacher at the front of the class writing, presenting and reading from the text book (EIA 2010). EIA encourages teachers to reduce the amount of presenting by the teacher, to facilitate student activity and talk. Table 36 examines this presenting by teachers for all classrooms, and this indicates success for EIA in that the treatment group shows a statistically significant reduction in this time in contrast to the control group which shows no significant change (and indeed it increases). Consequently, there is a statistically significant difference in differences reinforcing the importance of this improvement, with an effect size just below 0.2.

Table 36 Teacher talk presenting as a percentage of all teacher talk, all classrooms (N=163)

	Pre-test	Post-test	Difference
EiA (N=94)	48.9	42.6	-6.3*$^{.013}$
Non-EiA (N=69)	46.5	49.6	3.1^{ns}
Difference	2.4^{ns}	-7.0*$^{.024}$	**DD= -9.4***

2-sided p = .023, 1-sided p = .012, r = 0.18 Note: significant differences <u>in</u> the table are tested one-sided.

Primary classrooms reflect this improvement, though the changes in presenting are more complex, with the treatment group showing an improvement and the control group showing an increase in presenting. Neither of these changes are statistically significant, but because they are moving in opposite directions, produces a statistically significant difference in differences (again with an effect size just below 0.2, a small-to-medium effect size; Table 37).

Table 37 Teachers presenting as a percentage of all teacher talk, Primary (N=79)

	Pre-test	Post-test	Difference
EiA (N=45)	42.8	39.1	-3.7^{ns}
Non-EiA (N=34)	39,3	45,5	6,2^{ns}
Difference	3,5^{ns}	-6,4^{ns}	**DD= -9,9***

2-sided p = .099, 1-sided p = .05, r = .19 Note: significant differences <u>in</u> the table are tested one-sided.

The situation in the secondary classroom also indicates a statistically significant improvement in the teacher presenting (it was reduced) in the treatment group (Table 38), whereas in the control group shows little change, leading to a difference in differences that is not quite statistically significant. (The effect size is slightly lower than for primary classrooms, 0.17).

Table 38: Teachers' presenting as a percentage of all teacher talk, Secondary (N=84)

	Pre-test	Post-test	Difference
EiA (N=49)	54,5	45,9	-8,6*[.011]
Non-EiA (N=35)	53,4	53,6	0,2[ns]
Difference	1,1[ns]	-7,7[ns]	**DD= -8,8[ns]**

2-sided p = .123, 1-sided p = .062, r = .17 Note: significant differences <u>in</u> the table are tested one-sided.

Teachers' talk organising student activity (as percentage of teacher talk-time)

Although EIA encourages a reduction in teacher talk time it also seeks that the talk is productive, and time spent 'organising' the classroom is part of this. Table 39 gives the data on this organising for all classrooms observed, indicating that there was a statistically significant increase in organising time for the EIA treatment group, with no significant change for the control group, but the difference in differences is not statistically significant, though positive.

Table 39: Teachers organising activity as a percentage of total teacher talk, all classrooms (N=163)

	Pre-test	Post-test	Difference
EiA (N=94)	20,1	27,0	6,9**[.002]
Non-EiA (N=69)	19,2	22,4	3,2[ns]
Difference	0,9[ns]	4,6[.061]	**DD= 3,7[ns]**

2-sided p = .281, 1-sided p = .141, r = .08 Note: significant differences <u>in</u> the table are tested one-sided.

Looking at the two phases, EIA primary teachers show an improvement in organising talk (Table 40), unlike that of the control teachers, but again the difference in differences is not statistically significant, albeit positive. The picture for secondary teachers (Table 41) is similar to that for primary teachers.

Table 40: Teachers talk organising activity as a percentage of total teacher talk, Primary (N=79)

	Pre-test	Post-test	Difference
EiA (N=45)	26,1	32,7	6,6*[.027]
Non-EiA (N=34)	24,0	25,5	1,5[ns]
Difference	2,1[ns]	7,2*[.030]	**DD= 5,1[ns]**

2-sided p = .309, 1-sided p = .155, r = .11 Note: significant differences <u>in</u> the table are tested one-sided.

Table 41: Teachers talk organising activity as a percentage of total teacher talk, Secondary (N=84)

	Pre-test	Post-test	Difference
EiA (N=49)	14,6	21,7	7,1*[.012]
Non-EiA (N=35)	14,5	19,4	4,9[ns]
Difference	0,1[ns]	2,3[ns]	**DD= 2,2[ns]**

2-sided p = .631, 1-sided p = .316, r = .05 Note: significant differences <u>in</u> the table are tested one-sided.

3.5　Demographic effects: Gender and location effects on EL proficiency.

In this section Students' English Language Proficiency (GESE grade), is disaggregated by gender and by location (urban/rural/semi-urban) to determine the difference in differences and effect sizes. The sample overall was sufficient for the planned analysis across experimental conditions to detect an effect of 0.2, and therefore it was realised that sample sizes for disaggregated demographics would be necessarily lower, to the extent that effect sizes of 0.2 might not have been detected. However, there were some significant effect sizes in these smaller groups (e.g. male secondary students), and therefore these results are presented as a whole. Where there is a concern for the sample size this is indicated for particular analyses.

Students GESE disaggregated by Gender

Table 42 gives the distribution of the samples for the treatment and the control groups by phase and by gender. Table 43 gives the results for ELC (GESE).

Table 42: Gender Distribution, all students (N=1802)

EiA				Non-EiA			
1102				700			
Primary		Secondary		Primary		Secondary	
620		482		369		331	
Male	Female	Male	Female	Male	Female	Male	Female
267	353	163	319	163	206	155	176

Table 43: Student GESE by Gender (summary)

	Control			Treatment			Diff in Diff	Statistical significance ($p <0.05$) (1-sided)	Effect Size r
Variable	**Pre**	**Post**	**Diff**	**Pre**	**Post**	**Diff**			
Male student GESE grade	1.3	1.4	0.1	1.0	1.3	0.3	0.2	$p = .008$**	.09
Female student GESE grade	1.0	1.2	0.2	1.1	1.3	0.2	0.0	$p = .351^{ns}$	.01
Primary Male student GESE grade	0.5	0.7	0.2	0.6	0.8	0.2	0.0	$p = .487^{ns}$	.00
Primary Female student GESE grade	0.5	0.6	0.1	0.5	0.7	0.2	0.1	$p = .191^{ns}$	.04
Secondary Male student GESE grade	2.1	2.1	0.0	1.6	2.1	0.5	0.5	$p = .002$**	.16
Secondary Female student GESE grade	1.5	1.8	0.3	1.7	1.9	0.2	-0.1	$p = .159^{ns}$	.04

As Table 42 indicates, there are improvements from pre- to post-test for almost all treatment groups, and for most treatment groups (primary male, primary female, secondary female) this improvement is consistently 0.2 GESE grades (and as the tables below indicate these improvements

are statistically significant). For most control groups, there are similar statistically significant improvements of 0.1-0.2 GESE grades. Therefore, for most conditions, there is therefore a small difference-in-differences of 0 to 0.1 (as the tables below indicate these are not statistically significant).

The exception is for secondary male students (Table 48), where the treatment group performs comparably better than other treatment groups (a statistically significant increase in 0.5 GESE grade), whereas the control group has no difference from pre-test to post-test. These together mean that the difference-in-differences for male secondary students is larger than for other groups and is statistically significant, with a small-to-medium effect size (r=0.16). This results in secondary male students (control and treatment) having a statistically significant increase for male students overall (primary and secondary combined), although this is only of a small effect size (r~0.1), as Table 43 indicates. The details of these effects are detailed in the tables that follow.

Table 44Table 48 shows the details of the discussion above, with the highly significant difference in both the pre- and post-test improvements in both the control and treatment groups for male students overall, and a statistically significant difference in differences with small-medium effect size of 0.2. Although female students overall also show statistically significant improvements pre- to post-test, there is no difference in differences (Table 45).

Table 44: All Male students' GESE (N= 748)

	Pre-test	Post-test	Difference
EiA (N=430)	1.0	1.3	0.3***[<.001]
Non-EiA (N=318)	1.3	1.4	0.1**[.009]
Difference	-0,3**[.002]	-0,1[ns]	**DD =0,2***

2-sided p = .015, 1-sided p = .008, r = .09 Note: significant differences in the table are tested one-sided.

As noted above, Table 45 shows that, although there are statistically significant improvements for female students, pre to post-test in both control and treatment groups, there is no difference-in-differences.

Table 45: All Female students' GESE (N= 1054)

	Pre-test	Post-test	Difference
EiA (N=672)	1.1	1.3	0.2***[<.001]
Non-EiA (N=382)	1.0	1.2	0.2***[<.001]
Difference	0.1[ns]	0.1[ns]	**DD = 0.0 [ns]**

2-sided p = .702, 1-sided p = .351, r = .01 Note: significant differences in the table are tested one-sided.

Similarly, primary male students show a statistically significant improvement pre- to post-test but no difference in differences (Table 46). Their female counter parts do, however exhibit a difference –in-differences, but it is not statistically significant (Table 47).

Table 46 Primary Male GESE (N= 430)

	Pre-test	Post-test	Difference
EiA (N=267)	0.6	0.8	0.2***[<.001]
Non-EiA (N=163)	0.5	0.7	0.2***[<.001]
Difference	0.1[ns]	0.1[ns]	**DD = 0.0 [ns]**

2-sided p = .973, 1-sided p = .487, r = .00 Note: significant differences in the table are tested one-sided.

Table 47 Primary Female GESE (N= 559)

	Pre-test	Post-test	Difference
EiA (N=353)	0.5	0.7	0.2***[<.001]
Non-EiA (N=206)	0.5	0.6	0.1**[.005]
Difference	0.0[ns]	0.1[ns]	**DD = 0.1[ns]**

2-sided p = .382, 1-sided p = .191, r = .04 Note: significant differences in the table are tested one-sided.

Secondary students, as indicated earlier, show a gender effect, with male secondary student showing a statistically significant improvement from pre- to post-test (the control group does not improve), and a statistically high difference in differences with medium effect size of 0.5 (Table 48). Female secondary students (control and treatment) show an improvement from pre- to post-test, but do not show a significant difference in differences (Table 49).

Table 48: Secondary Male GESE (N= 318)

	Pre-test	Post-test	Difference
EiA (N=163)	1.6	2.1	0.5***[<.001]
Non-EiA (N=155)	2.1	2.1	0.0[ns]
Difference	-0.5**[.006]	0.0[ns]	**DD = 0.5****

2-sided p = .003, 1-sided p = .002, r = .16 Note: significant differences in the table are tested one-sided.

Table 49: Secondary Female GESE (N= 495)

	Pre-test	Post-test	Difference
EiA (N=319)	1.7	1.9	0.2**[.006]
Non-EiA (N=176)	1.5	1.8	0.3***[<.001]
Difference	0.2[ns]	0.1[ns]	**DD = -0.1[ns]**

2-sided p = .318, 1-sided p = .159, r = .04 Note: significant differences in the table are tested one-sided.

Students GESE disaggregated by Locations

It evident from Table 50 that most of the students are in rural locations (1060 with groups ranging from 175 to 429), with semi-urban being the next largest location (539 with groups from 76-173).[30]

---

[30] These proportions approximate those in the national school population

There are relatively few from urban locations (203), and within these, some very small groups (18-97). These latter groups are likely to suffer from sample effects, even though they may exhibit statistically significant differences. Because of these limitations of a highly uneven distribution of students across locations and the small or very small numbers of students in some locations, the findings for location should be considered as somewhat tentative.

Table 50: Sample distribution by location (N=1802)

EiA						Non-EiA					
1102						700					
Primary			Secondary			Primary			Secondary		
620			482			369			331		
Urban	Semi-urban	Rural	Urban	Semi-urban	Rural	Urban	Semi-urban	Rural	Urban	Semi-urban	Rural
18	173	429	97	128	257	32	162	175	56	76	199

The summary results for ELC assessment by location are shown in Table 51.

Table 51: GESE grade by location

	Control			Treatment			Diff in Diff	Statistical significance ($p < 0.05$) (1-sided)	Effect Size r
Variable	**Pre**	**Post**	**Diff**	**Pre**	**Post**	**Diff**			
Urban	1.44	1.35	-0.09	1.84	2,14	0.30	0.39	$p = .002**$	.20
Semi-urban	1.04	1.27	0.23	1.26	1.38	0.12	-0.11	$p = .156^{ns}$	.04
Rural	1.05	1.26	0.21	0.80	1.09	0.29	0.08	$p = .088^{ns}$	.04
Primary urban	0.63	0.44	-0.19	0.50	1.33	0.83	1.02	$p < .001***$	.59
Primary semi-urban	0.48	0.62	0.14	0.50	0.77	0.27	0.13	$p = .068^{ns}$	.08
Primary rural	0.47	0.75	0.28	0.54	0.70	0.16	-0.12	$p = .041*$	.07
Secondary urban	1.91	1.88	-0.03	2.09	2.29	0.20	0.23	$p = .087^{ns}$	.11
Secondary semi-urban	2.24	2.66	0.42	2.27	2.20	-0.07	-0.49	$p = .026*$	.14
Secondary rural	1.56	1.70	0.14	1.24	1.74	0.50	0.36	$p < .001***$	.16

The individual analyses for each of these groups are given in Table 52 to
Table 60. Almost all treatment groups show an improvement at post-test assessment (apart from secondary semi-urban).

As might be expected, urban students generally do better than students from other locations (the pre-test results are indeed higher, as might be expected) for both control and treatment groups.

*Urban students overall (Table 52) show a statistically significant difference-in-differences, with a small-medium effect size (0.2). Primary urban students in particular show a large difference-in-differences, of (1.02) more than an entire GESE grade, which is highly statistically significant and of a large effect size (0.59). However, this is the smallest sample of all (50 students) and this result should be therefore be treated with some caution (*Table 55*).*

Secondary rural students go against this trend with the treatment group showing the highest improvement pre- to post-test, which is highly statistically significant, producing a significant difference-in-differences but a small effect size (0.16). (Rural students overall show significant improvements in pre- to post-test assessment, but as this occurs for both control and treatment groups the differences-in-differences is not significant; see Table 54).

The tables that follow give all the detailed analysis for each of the location sub-groups in each phase.

Table 52: All Urban students GESE (N=203)

	Pre-test	Post-test	Difference
EiA (N=115)	1.84	2.14	$0.30**^{.003}$
Non-EiA (N=88)	1,44	1.35	-0.09^{ns}
Difference	$0.40^{.055}$	$0.79***^{<.001}$	**DD= 0,39****

2-sided p = .004, 1-sided *p* = .002, *r* = .20 Note: significant differences <u>in</u> the table are tested one-sided.

Table 53: All Semi-urban GESE (N= 539)

	Pre-test	Post-test	Difference
EiA (N=301)	1.26	1.38	0.12^{ns}
Non-EiA (N=238)	1.04	1,27	$0,23***^{<.001}$
Difference	$0.22^{.052}$	0.11^{ns}	**DD= -0.11^{ns}**

2-sided *p* = .311, 1-sided *p* = .156, *r* = .04 Note: significant differences <u>in</u> the table are tested one-sided.

Table 54: All Rural GESE (N= 1060)

	Pre-test	Post-test	Difference
EiA (N=686)	0.80	1.09	$0.29***^{<.001}$
Non-EiA (N=374)	1.05	1.26	$0.21***^{<.001}$
Difference	$-0.25***^{<.001}$	$0.17**^{.005}$	**DD= 0.08^{ns}**

2-sided *p* = .175, 1-sided *p* = .088, *r* =.04 Note: significant differences <u>in</u> the table are tested one-sided.

Table 55: Primary urban GESE (N=50)

	Pre-test	Post-test	Difference
EiA (N=18)	0.50	1.33	$0.83***^{<.001}$
Non-EiA (N=32)	0.63	0.44	$-0.19*^{.028}$
Difference	-0.13^{ns}	$0.89***^{<.001}$	**DD= 1.02*****

NB. **Small sample size**. 2-sided *p* <.001, 1-sided <.001, *r* **= .59** Note: significant differences <u>in</u> the table are tested one-sided.

Table 56: Primary semi-urban GESE (N= 335))

	Pre-test	**Post-test**	**Difference**
EiA (N=173)	0.50	0.77	0.27***$^{<.001}$
Non-EiA (N=162)	0.48	0.62	0.14**$^{.009}$
Difference	0.02^{ns}	0.15*$^{.050}$	**DD= 0.13^{ns}**

2-sided *p* = .136, 1-sided *p* = .068, *r* = .08 Note: significant differences <u>in</u> the table are tested one-sided.

Table 57: Primary rural GESE (N= 604)

	Pre-test	**Post-test**	**Difference**
EiA (N=429)	0.54	0.70	0.16***$^{<.001}$
Non-EiA (N=175)	0.47	0.75	0.28***$^{<.001}$
Difference	0.07^{ns}	-0.05^{ns}	**DD= -0.12***

2-sided *p* = .081, 1-sided *p* = .041, *r* =.07 Note: significant differences <u>in</u> the table are tested one-sided.

Table 58: Secondary urban GESE (N=153)

	Pre-test	**Post-test**	**Difference**
EiA (N=97)	2.09	2.29	0.20*$^{.045}$
Non-EiA (N=56)	1.91	1.88	-0.03^{ns}
Difference	0.18^{ns}	0.41*$^{.039}$	**DD= 0.23^{ns}**

2-sided *p* = .173, 1-sided *p* = .087, *r* = .11 Note: significant differences <u>in</u> the table are tested one-sided.

Table 59: Secondary semi-urban GESE (N= 204)

	Pre-test	**Post-test**	**Difference**
EiA (N=128)	2.27	2.20	-0.07^{ns}
Non-EiA (N=76)	2.24	2.66	0.42***$^{<.001}$
Difference	0.03^{ns}	-0.46*$^{.039}$	**DD= -0.49***

2-sided *p* = .052, 1-sided *p* = .026, *r* = .14 Note: significant differences <u>in</u> the table are tested one-sided.

Table 60: Secondary rural GESE (N= 458)

	Pre-test	**Post-test**	**Difference**
EiA (N=257)	1.24	1.74	0.50***$^{<.001}$
Non-EiA (N=199)	1.56	1.70	0.14*$^{.039}$
Difference	-0.32**$^{.003}$	0.04^{ns}	**DD= 0.36***

2-sided *p* <.001, 1-sided *p* <.001, *r* = .16 Note: significant differences <u>in</u> the table are tested one-sided.

4.   Summary of Findings

4.1   Overview

The study was a requirement of the 2014 Annual Review, strongly supported by DFID Bangladesh. This was a quasi-experimental study, where a control group of schools is compared with a treatment group of schools with measures taken pre-intervention and post-intervention. (In addition, it is being used to give the relevant output and outcome targets of the Logical Framework [logframe]). Stratified random samples were used to select Upazilas and schools proportionately from each division, with selected schools from each sampled Upazila being randomly assigned to control or treatment groups. Sampling was carried out separately for primary and secondary schools to avoid bias.

English Language teachers from treatment schools took part in a Programme of school-based teacher development (SBTD). Teachers from control schools did not participate in the Programme until after the completion of the study. The SBTD Programme was driven by teachers: observing Communicative Language Teaching (CLT) techniques demonstrated through video clips of authentic teachers and students in their classrooms; then practicing these techniques and introducing them in their own teaching. Teachers were supported to do this by: four local face-to-face cluster meetings intended to run over one school year; working together with another teacher from their school; by their head teacher; and by audio-visual resources available offline, on a memory card, accessed on the teachers' own mobile phone.

In fact, teachers began the treatment (the SBTD Programme) between August-to-November 2015 and ended between July-to-October 2016 [9-14 months]. At the time of the post-test (May 2016), treatment teachers were only half-way through the SBTD Programme (having completed two of four teacher development meetings). Although on paper, they had been in the Programme for six-to-nine months, in reality the only effective 'teaching time' available for the introduction of new classroom activities was limited to around two-to-three months, from February/March to April/May 2016[31]. This gave limited time for teachers to integrate EIA classroom activities into their regular practice and an unrealistically short period for these changes to translate into measurable improvements in learning outcomes, and should be viewed as only a 'partial' treatment.

In total 120 schools (and hence 240 teachers and 2,400 students) were planned to be involved in the studies (Table 2 gives a summary of the planned sample). In order to allow for student attrition, class changes and drop-out, the pre-test sample had 50% more students than the minimum sample size suggested by power analysis prior to the study. (This did not significantly increase the analytical power or the cost of the study, but made the sample more robust to the effects of any reduction in the sample as indicated.) Table 6 to Table 9 gave the actual samples, as part of the findings. The analysis calculated the relative changes (the *difference*) in student learning outcomes over the period from the pre-intervention to the post-intervention, and compare these differences between control and treatment groups (using a *difference-in-differences* measure). The study also examined in what ways (if any) teacher's classroom practices changed.

---

[31] See section 2.5 Timeline

4.2 Key Findings

Classroom Practices

There were several statistically significant experimental effects on classroom practice, with effect sizes ranging from small (r=0.1) to medium (r=0.27). Several of these have statistical significance only in one phase (i.e. primary or secondary lesson observations), whilst across all observations (primary and secondary) similar positive difference-in-differences were seen, but without achieving statistical significance.

Statistically Significant Experimental Effects on CP[32]

1. **Increased student talk in English** as percentage of student talk time
   a. In secondary lesson observations[33] there was a statistically significant increase in students talk in English in treatment schools (p=0.04), with no statistically significant difference in control schools. The DID (+15% points) was statistically significant (p=0.04), with a small-to-medium effect size (r=0.19).
   b. Across all lesson observations (primary and secondary combined),[34] there was a statistically significant (p=0.03) increase of 6% points (from 82% to 88%) in students use of English language in treatment schools, whilst there was no statistically significant difference in control schools. The DID (+4.5% points) across all lesson observations was positive, but not statistically significant.
2. **Increased student-student talk** as percentage of all student talk:
   a. In primary lesson observations[35] there was a dramatic four-fold increase (from 4% to 16%) in student-student talk in treatment schools, which was highly statistically significant (p<0.001). There was no statistically significant difference between pre- and post-test results in control schools. The difference in differences (14% points) was large and was highly statistically significant (p=0.006). There was a medium effect size (r=0.27)
   b. Across all observations (primary and secondary combined)[36], student-student talk in treatment schools increased by 4% points (from 11% to 15%); the increase was almost statistically significant (p=0.08). There was no statistically significant difference between pre- and post-test in control schools. The DID (+3.6% points) across all lesson observations was positive but not statistically significant.
3. **Increased teachers' talk in English**, as a percentage of teacher talk time
   a. In Secondary lesson observations[37] there was a substantial increase in teachers talk in English of 23% points (from 64% to 86%) in treatment schools. The difference (23%) was highly statistically significant (p<0.001). For control schools, there was no statistically significant difference. The difference in differences (16% points) was large and statistically significant (p=0.18). The effect size was small-to-medium (r=0.21).
   b. Across all observations (primary and secondary combined),[38] teachers' talk in English increased 11% points (from 63% to 74%) in control schools, but the increase was greater (18% points) in treatment schools, reaching 87% at post-test. The DID (+7% points) across all lesson observations was positive but not statistically significant.

---

[32] Some of the percentages are rounded to the nearest whole number and where appropriate, the tables referred to for detailed data.
[33] Table 25: Students talk in English, Secondary (N= 84)
[34] Table 23: Student's talk in English, all classrooms (N=163)
[35] Table 28: Student-Student talk as percentage of all student talk, Primary (N=79)
[36] Table 27: student-student talk as percentage of all student talk, all classrooms (N=163)
[37] Table 35 Teachers Talk in English, Secondary (N=84)
[38] Table 33 Teachers talk in English, all classrooms (N=163)

4. **Decreased teacher talk 'presenting'** to students
   a. Across all lesson observations (primary and secondary combined),[39] in treatment schools there was a 6% points reduction (from 49% to 43%) in teacher talk-time given to 'presenting'. The reduction was statistically significant (p=0.01). For control schools, there was no statistically significant difference between the pre- and post-tests. The difference in differences (-10% points) was negative (the treatment seeks to reduce teachers 'presenting') and was statistically significant (p=0.01) with a small-to-medium effect size (r=0.18).

Other Statistically Significant Findings on Classroom Practice

There were also statistically significant changes in two further indicators of classroom practice, although these were not associated with statistically significant difference-in-differences:

5. **Decreased teacher talk-time** as percentage of lesson time
   a. Across all lesson observations (primary and secondary combined),[40] in treatment schools, there was a reduction in teachers talk time from 53% to 49%, which was statistically significant (p=0.03). There was no statistically significant change in control schools. The difference-in-differences (-2.5% points) was negative (the treatment seeks to reduce teachers talk time) but not statistically significant.
6. **Increased teacher talk-time organising student activity**, as a percentage of teacher talk-time
   a. Overall (primary and secondary combined)[41] teachers organising student activity increased 7% points (from 20% to 27% of teacher talk), with the increase being highly statistically significant (p=0.002). The DID was 4% points.

Other Findings on Classroom Practice

There was one further indicator of classroom practice where statistically significant change was *not* found:

7. **Student talk-time** as percentage of lesson time
   a. Across all lesson observations (primary and secondary combined),[42] student talk-time was high at pre-test (20% in control and treatment). There were small increases at post-test, but these were not statistically significant for control or treatment. There was a small positive DID (1.1%), but this was not statistically significant.
   b. The pre-test findings contrasted with earlier baseline studies which showed very low levels of student talk (EIA, 2011), suggesting teachers had generally become aware of the importance of encouraging student talk in English Language lessons. In Programme terms this may be considered as institutionalisation, or in experimental terms, as contamination.

Student Learning Outcomes: English Language Proficiency

- Student English Language Proficiency (primary and secondary combined),[43] improved more for treatment students (+0.3 GESE grades) than control (+0.2 GESE grades). These improvements were highly statistically significant (p<0.001) for both groups.
- The difference in differences (0.1 GESE grade) was positive but not statistically significant.

---

[39] Table 36 Teacher talk presenting as a percentage of all teacher talk, all classrooms (N=163)**Error! Reference source not found.**
[40] Table 30: Teachers talk as percentage of lesson time, all classrooms (N=163)
[41] Table 39: Teachers organising activity as a percentage of total teacher talk, all classrooms (N=163)
[42] Table 19: Student Talk time as a percentage of lesson time, summary experimental results
[43] Table 16: ELC (GESE) All students (N=1802)

- When disaggregated, the difference in differences for secondary students[44] was positive (+0.1 GESE grades) but not statistically significant; there was no difference in differences for Primary students.[45]

However, when disaggregated a number of positive experimental effects were found, below.

Statistically Significant Experimental Effects on Students English Language Proficiency

8. Male Students
    a. There was an increase of 0.3 GESE grades (from 1.0 to 1.3) for male treatment students (primary and secondary combined) which was highly statistically significant ($p<0.001$). There was an increase of 0.1 GESE grades (from 1.3 to 1.4) for male control students, which was also highly statistically significant ($p=0.009$). The difference in difference of 0.2 GESE grades was highly statistically significant ($p=0.008$), and there was a small effect size ($r=0.1$).

    b. Male Secondary Students[46] There was a large increase of 0.5 GESE grades (from 1.6 to 2.1) for treatment schools, that was highly statistically significant ($p<0.001$). For control schools, there was no difference at all between pre- and post-test GESE scores (which remained constant at 2.1). The difference in differences (0.5 GESE grades) was highly statistically significant ($p=0.002$), and there was a small effect size ($r=0.16$).

Two factors combine to make the DID for male secondary students larger than for other treatment groups:

> Firstly, in treatment schools, male and female primary students and female secondary students all evidenced the same increase in English Language Proficiency: 0.2 GESE grades. But Secondary male treatment students showed a larger improvement of 0.5 GESE grades.

> Secondly, in control schools, male and female primary students and female secondary students all evidenced the similar increases in English Language Proficiency of 0.1-0.2 GESE grades. But secondary male control students were the only control group to show *no improvement* at post-test.

These two factors combined to make a larger difference-in-difference for secondary males than for other groups and also pushed male students overall (primary and secondary combined) into a positive experimental effect. But it should be emphasised that treatment groups for male and female primary students and female secondary students all improved between pre- and post-test by the same extent (0.2 GESE grades).

9. Urban Students
    a. There was an increase of 0.3 GESE grades (from 1.8 to 2.1) for urban treatment schools (Primary and Secondary combined),[47] that was highly statistically significant ($p=0.003$). There was no statistically significant difference for control schools. The difference in differences was almost half a GESE grade (0.39) and was highly statistically significant ($p=0.002$). There was a small-to-medium effect size for urban students ($r=0.2$).

---

[44] Table 18: Secondary students GESE (N= 813)
[45] Table 17 Primary students GESE (N= 989)
[46] Table 48: Secondary Male GESE (N= 318)
[47] Table 52: All Urban students GESE (N=203)

40

    b. There was a large increase of 0.83 GESE grades (from 0.5 to 1.33) for urban primary treatment students[48], which was highly statistically significant ($p<0.001$). There was a small but statistically significant ($p=0.28$) decrease in GESE grades for control students (from 0.63 to 0.44). The increase in treatment and decrease in control led to a difference in differences of 1.02 GESE grades, which was highly statistically significant ($p<0.001$), with a large effect size ($r=0.59$).

10. Rural students
    a. There was an increase of 0.5 GESE grades (from 1.24 to 1.74) for secondary rural treatment students[49]. This increase was highly statistically significant ($p<0.001$). There was a smaller increase of 0.14 GESE grades (from 1.56 to 1.7) for control students, which was statistically significant ($p<0.04$). The difference in differences of 0.36 GESE grades for secondary rural students was highly statistically significant ($p<0.001$), with a small effect size ($r=0.16$).

Most other gender or location disaggregated groups showed highly statistically significant improvements in English language proficiency pre- to post-test, but the differences-in-differences between control and treatment were not statistically significant.

There were two groups where (just) statistically significant differences-in-difference were found showing control improving more than treatment:

11. Semi-urban students (secondary)[50]: increase of 0.42 GESE grades for control, highly statistically significant (<0.001) but no improvement in treatment, consequently a difference in differences of -0.49, statistically significant ($p=0.03$), and a small effect size ($r=0.14$).
12. Rural students (primary)[51]: highly statistically significant improvements for control and treatment groups, but difference in differences of -0.12 GESE grades (negative), which was significant ($p=0.04$), but with an effect size less-than-small ($r=0.07$).

Semi-urban secondary students had the highest pre-test scores of any group (Table 51) both for control (average GESE score 2.24) and treatment (average GESE score 2.27) students. As noted previously, gains at higher levels take longer to achieve and due to nationwide protests, schools were only halfway through treatment at post-test; it is therefore likely that if post-test had been at the end of the treatment, greater positive DID would have been seen for semi-urban secondary students.

Primary rural students were the group most affected by contamination. The qualitative companion study to this report found numerous examples of contamination amongst primary schools (from three out of four primary schools in the study) with some indication that contamination may be more pro-active in rural areas, where head-teachers actively share anything learned from training (ideas, practices or resources) between local schools. The primary rural sample was also most

---

48

Table 55: Primary urban GESE (N=50)
49

Table 60: Secondary rural GESE (N= 458)
[50] Table 59: Secondary semi-urban GESE (N= 204)
[51] Table 57: Primary rural GESE (N= 604)

affected by schools being withdrawn from the study due to contamination (Table 7), resulting in the treatment group being more than twice as large as the control group (Table 57).

5.   Conclusions

The study set out to investigate if EIA in the institutionalised phase (IV) had met its logframe targets, and to provide more robust evidence for the efficacy of EIA (through a quasi-experimental approach), and to enable this evidence to be used internationally. It set out to do this in a social and political situation of great unrest, and thus with implementation difficulties for EIA that led to a relatively short treatment time, compared with what was already a rather limited planned time. Both mitigate against the achievement of the aims of the study.

The provision of evidence of achievement of the logframe targets was particularly strengthened by the using pre- and post-test model, where the same students were assessed for ELC and the same classrooms observed each time, compared to previous cohort studies that chose different samples at pre- and post-test assessments and observations. This study showed that on the two indicators EIA has largely achieved its targets:

- *Outcome indicator OM1a: ELC for primary and secondary students has improved*.
  It is evident that for both primary and secondary students there are statistically significant improvements in the GESE scores from pre- to post-test, and that for criteria of achievement, EIA exceeded the lower set of criteria ($\geqslant$ GESE 1 for primary and $\geqslant$ GESE 2 for secondary), but only partially achieved (2%) the 5% improvement target at the higher set ($\geqslant$ GESE 2 for primary and $\geqslant$ GESE 3 for secondary).
- *Output indicators O1b & 3b: improvement in classroom practices of teachers*.
  Both primary and secondary teachers achieved the targets and both substantially exceeded the targets for the percentage of student talk in English.

On the second purpose, to provide evidence of the effectiveness of EIA compared to a control group, the evidence is complex and a summary is set out below. Statistically significant experimental effects on classroom practices and learning outcomes (students English Language Proficiency) were found for some groups in the study, with a range of effect sizes from small (r=0.1) to large (r=0.6). Most effect sizes were in the small (0.1) to medium (0.3) range. As a result of the EIA treatment, there were statistically significant experimental effects (DID) on:

1. Classroom Practices.
   a. All **teachers spent less time 'presenting'** to students. (A statistically significant difference-in-difference of -9% points, with a small-to-medium effect size of r=0.2)
   b. In primary lessons, **there was more student-student talk**. (A statistically significant difference-in-difference of 15% points, with a medium effect size of r=0.3).
   c. Secondary **teachers used more spoken English**. (A statistically significant difference-in-difference of 16% points, with a small-to-medium effect size of r=0.2)
   d. Secondary **students used more spoken English** (A statistically significant difference-in-difference of 15% points, with a small-to-medium effect size of r=0.2)
2. Students' English Language Competence (GESE grade)
   a. **Male students (primary, secondary, all locations) improved their English Language Competence** (A statistically significant difference-in-difference of 0.2 GESE grades with a small effect size of r=0.1)

b. **Urban students (primary, secondary, boys and girls) improved their English Language Competence**. (A statistically significant difference-in-difference of 0.3 GESE grades, with small to medium effect size of *r*=0.2).

c. **Rural students (secondary only, boys and girls)** improved their English Language Competence. (A statistically significant difference-in-difference of 0.36 GESE grades, with small-to-medium effect size of *r*=0.2.

For two groups (rural primary students and semi-urban secondary students), the control students improved their GESE grades more than treatment students. The DIDs, statistical significance and effect sizes were generally smaller than the positive experimental effects, but still significant.

The following key conclusions can be drawn from the analysis:

I. There are statistically significant improvements in classroom practice and learning outcomes, but these are seen mostly for either primary or secondary phases, rather than across the whole-study populations (of students and teachers).

II. These experimental effects are mostly of a small (*r*=0.1) to medium (*r*=0.27) effect size, although one is of large effect size (*r*=0.6).

III. There are wider statistically significant improvements beyond the experimental effects, often associated with positive though not statistically significant differences-in-difference.

We identify two factors that have limited the extent to which statistically significant experimental effects were found:

IV. Treatment schools received only partial exposure to the full treatment:

With the post-test having to take place immediately after the second of four teacher development meetings scheduled over the year, teachers had only a few months of teaching time to develop improved classroom practices and for these to translate into student learning outcomes (see section 2.5).

V. Control schools also received some exposure to treatment, through contamination:

- from experimental treatment within their Upazilas

There is some evidence from the qualitative study that teachers in control schools were aware of, and in some cases emulating, practices taking place in treatment schools. The teachers from control and treatment schools took part together in meetings across the Upazilas, including Needs-Based Sub-Cluster (NBSC) Meetings and Teachers Support Network (TSN) meetings, whilst head teachers were brought together at Head Teacher meetings. In all of these forums, teachers from treatment and control schools will have had the opportunity to share their understandings and practices. Additionally, local Education Officers, who participate in the organisation and oversight of treatment, organise and deliver most of the Upazila level events and also visit schools, working across control and treatment settings, and hence had opportunities to disseminate treatment approaches.

- from national institutionalisation of treatment.

EIA has been partially integrated into the national Programme of Subject Based Training for English language teachers, as well as the national textbook series and the Teachers' Editions

of these (with lesson guidance), as well as the associated 'Curriculum Dissemination Training'. EIA digital materials are also made available through the governments Multi-Media Classrooms (MMC) Programme. It has not been possible to track how many teachers (control or treatment) have been exposed to aspects of the treatment through participation in these various Programmes, or when, during the study.

- from prior exposure of Upazilas to treatment.

  Whilst none of the schools in the study had taken part in EIA previously, almost half of the Upazilas (24 of 55, 44%) had been involved in previous cohorts of EIA. Some of these (11 Upazilas) had participated in EIA for just one year previously, but others (7 Upazilas) had two previous years of participation in EIA and almost as many (6 Upazilas) had seven years of exposure to EIA. The greater time an Upazila has had exposure to the treatment (EIA), the more likely it is that the practices and ideas promoted through the treatment will have spread (via teachers, teacher-facilitators and education officers), from school-to-school within the Upazila, and even to schools with no formal exposure to the treatment.

Despite these limitations, this study does show a number of statistically significant experimental effects, both in terms of more communicative classroom practices and improved learning outcomes (students English Language Competency). Had it been possible to conduct the post-test at the end of the teacher development Programme, as initially planned, the authors consider it reasonable to expect that these changes would have become more widely distributed across the treatment populations and with greater effect sizes.

6. References

Boissiere, M. (2004) *Determinants of Primary Education Outcomes in Developing Countries: Background Paper for the Evaluation of the World Bank's Support to Primary Education*. Operations Evaluation Department. Washington DC.

English in Action (EIA) (2009) An Observation Study of English Lessons in Primary and Secondary Schools in Bangladesh. Baseline Study 3, English In Action, Dhaka, Bangladesh: English In Action.

English in Action (EIA) (2012) *English Proficiency Assessments of Primary and Secondary Teachers and Students Participating in English in Action (Study 3a).* Dhaka, Bangladesh: EIA.

English in Action [EIA] (2011) *The Classroom Practices of Primary and Secondary School Teachers Participating in English in Action*. Dhaka, Bangladesh: EIA.

Hattie, J. (2013) *Visible Learning: A Synthesis of Over 800 Meta-Analyses Relating to Achievement*. Abingdon, UK: Taylor & Francis.

Krippendorff, K. (2004) Reliability in content analysis: Some common misconceptions and recommendations. *Human Communication Research*, 30(3):411–433.

Trinity College London (2007) *Relating the Trinity College London GESE and ISE examinations to the Common European Framework of Reference – project summary*. London: Trinity College London.

Trinity College London (2009) *Graded Examinations in Spoken English (GESE). Syllabus from 2010.* London: Trinity College London.

Trinity College London (2013) *Examination Information: Graded Examinations in Spoken English (GESE).* London: Trinity College London.

UNICEF (2010) *A case for geographic targeting and basic social service to mitigate inequalities in Bangladesh*. Technical report, UNICEF.

Westbrook, Jo *et al*. (2013) *Pedagogy, Curriculum, Teaching Practices and Teacher Education in Developing Countries. Final Report. Education Rigorous Literature Review*. London: Department for International Development.

7.   Appendices

7.1 Appendix 1: Classroom Observation Instrument and Guidance Notes

Teacher ID: _____

School ID: _____

(internal use)

Researcher information

Researcher name		Observation date	

School information

School name			
Upazila/Thana		Division	

Teacher information

Teacher name		Gender:	Male / Female
Age:		Highest qualification:	

Lesson information

Class:		Lesson no:			
Lesson start time:		Lesson finish time:		Duration of lesson (mins):	
Enrolment:	Total no. of students:	No. of boys:		No. of girls:	
Attendance:	Total no. of students in class:	No. of boys in class:		No. of girls in class:	

On each minute, identify what is happening at that moment of observation. Write E (for 'English') or B (for 'Bangla') in the appropriate box to show whether that moment of the lesson is being taught in English or Bangla.

- Enter 'E' or 'B' in <u>one</u> of the columns under:
  - o 'teacher is speaking' (if the teacher is speaking)
    <u>or</u>
  - o 'student(s) is speaking' (if a student(s) is talking)
    <u>or</u>
  - o 'students are' (if the student is carrying out an activity).

In addition, if visual materials are being used at the moment of observation, indicate what is being used in the '*Visual materials being used'* column. Use the following letters to indicate the kind of materials being used:

P = EIA poster or wall chart

C = EIA cards (flash-cards) (used with Primary only)

O = other visual aids (e.g. teachers own, perhaps a magazine, newspaper)

If neither the teacher nor the students are speaking, and the students are not reading, writing or listening to audio either, use the 'Other activity' column to indicate what other kind of activity is taking place. Make sure you do this for the duration of the entire class.

Further to this, remember to obtain consent from the teacher via the consent form. Ensure that they sign, date and record the time on 2 forms (the teacher should keep a copy of the form; and you should return the other to EIA project staff at the end of the fieldwork).

Time	The **teacher** is speaking. They are:				The **student(s)** are speaking.				Students are:			Enter P, C or OM	Enter other activity (if appropriate)
minute	Presenting	Organizing	Asking questions	Giving feedback	On their own (single)	In pairs	In groups	Chorusing	Reading	Writing	Listening to audio	Visual materials being used	Other activity
1													
2													
3													
4													
5													
6													
7													
8													
9													
10													
11													

Insert the letter E or B once per row in these columns, if appropriate

49

Time	The **teacher** is speaking. They are:				The **student(s)** are speaking:				Students are:			Enter P, C or O if VM	Enter other activity (if appropriate)
					Insert the letter E or B once per row in these columns, if appropriate								
minute	Presenting	Organizing	Asking questions	Giving feedback	On their own (single)	In pairs	In groups	Chorusing	Reading	Writing	Listening to audio	Visual materials being used	Other activity
12													
13													
14													
15													
16													
17													
18													
19													
20													
21													
22													

Time	The **teacher** is speaking. They are:				The **student(s)** are speaking:				Students are:			Enter P, C or O if VM	Enter other activity (if appropriate)
			Insert the letter E or B once per row in these columns, if appropriate										
minute	Presenting	Organizing	Asking questions	Giving feedback	On their own (single)	In pairs	In groups	Chorusing	Reading	Writing	Listening to audio	Visual materials being used	Other activity
23													
24													
25													
26													
27													
28													
29													
30													
31													
32													

Continuation sheet for 1 hr lessons

	The **teacher** is speaking. They are:				The **student(s)** are speaking:		Students are:					Enter P, C or O if VM	Enter other activity (if appropriate)
Time													
minute	Presenting	Organizing	Asking questions	Giving feedback	On their own (single)	In pairs	In groups	Chorusing	Reading	Writing	Listening to audio	Visual materials being used	Other activity
33													
34													
35													
36													
37													
38													
39													
40													
41													

Insert the letter E or B once per row in these columns, if appropriate

Time	The **teacher** is speaking. They are:				The **student(s)** are speaking:				Students are:			Enter P, C or O if VM	Enter other activity (if appropriate)
minute	Presenting	Organizing	Asking questions	Giving feedback	On their own (single)	In pairs	In groups	Chorusing	Reading	Writing	Listening to audio	Visual materials being used	Other activity
42													
43													
44													
45													
46													
47													
48													
49													
50													
51													

Insert the letter E or B once per row in these columns, if appropriate

	Insert the letter E or B once per row in these columns, if appropriate											Enter P, C or O if VM	Enter other activity (if appropriate)
Time	The **teacher** is speaking. They are:				The **student(s)** are speaking:				**Students** are:				
minute	Presenting	Organizing	Asking questions	Giving feedback	On their own (single)	In pairs	In groups	Chorusing	Reading	Writing	Listening to audio	Visual materials being used	Other activity
52													
53													
54													
55													
56													
57													
58													
59													
60													

Remember to obtain consent from the teacher via the consent form. Ensure they sign, date and record the time on 2 forms (the teacher should keep a copy of the form; and you should return the other to EIA project staff at the end of the fieldwork).

DEFINITIONS

The teacher is speaking
Presenting

The teacher is giving information to the students. They may be describing, explaining or narrating, whether from the textbook or from their own knowledge, or from any other source. Students are expected to listen to the information. Examples include:

- T is reading from a book.
- T is modelling the target language (past tense): *Yesterday I went to the market.*
- T points to poster. *Look at the picture.* T points to the tree. *The bird is in the tree.*

Organizing

The teacher is telling the students what to do. The students are expected not only to listen, but also to do something according to the teacher's directions. Examples include:

- *Read for five minutes and answer then answer the questions.*

- *We are going to listen to an audio about Mother Teresa.*

- *Listen to the audio and answer the two questions on the board.*

- *Stand up… sit down… make groups.*

- *Remember you don't need to read every word.*

- *Stand up when you have finished.*

Asking questions

The teacher is asking questions. The students are expected to respond verbally (as opposed to organizing, when the students respond non-verbally). Examples include:

- *What is the Bangla for 'magazine'?*
- *Can you describe the diagram?*
- *What do you think the girl is going to do?*
- *Who are the people in the picture?*

Giving feedback

The teacher is responding to something the students have said or done, and evaluating or commenting on it. Examples include:

- *Yes, that's correct.*
- *Not quite right. You need to use past tense.*
- *Well done, students.*
- *Oh, your picture looks very nice. But where is the river?*
- *S: He is catching the bus. T: He? S: Oh, she is catching…*

The student(s) are speaking

On their own (Single)

One student is speaking at this particular moment. The student may be talking to the teacher or with another student, or s/he may be reading aloud.

In pairs

All of the students are talking to each other in pairs.

In groups

All of the students are talking to each other in groups.

Chorusing

All of the class is speaking in chorus at the same time. This may be in response to the teacher's questions, or reading in chorus.

The students are reading
All or most of the students are reading something quietly. (If they are reading aloud, enter the activity under 'Student(s) are speaking')

The students are writing
All or most of the students are writing something quietly. (If they are discussing a writing task in pairs or groups, enter the activity under 'Student(s) are speaking')

The students are listening to audio
The teacher is playing an audio resource and students are listening.

Other activity
This could be any activity taking place in the classroom which does not fit into one of the categories above. For example:

1. Teacher is preparing learning materials.
2. Teacher is using the blackboard.
3. Teacher is checking students' work.
4. Teacher is doing administrative work.
5. Teacher is asking students to bring things from outside.
6. Students are getting in to pairs or groups.

7.2   Appendix 2: Calculation and Use of Effect Sizes

This report uses effect sizes, to enable comparison with other studies, where different outcome indicators have been used. Such comparisons between studies are known as 'Meta-analyses'. Effect sizes have been calculated using Pearsons' r[52], using a specific SPSS plug-in module designed for this purpose[53].

It is common to interpret Effect sizes, so that[54]:
- If r = 0.1, the effect size is 'small'
- If r = 0.3, the effect size is 'medium' and
- If r = 0.5, the effect size is 'large'.

But what does it mean, to say the effect size is small, medium or large? In the paragraphs below, we illustrate these effect sizes[55], by showing how many of the control group would be below average for the treatment group, at each effect size:

A small effect size (r=0.1) is one where the difference is not very substantial. For example, if there were 100 people in the control group:
- 42 of them (less than half) would score as well as an average person in the treatment group;
- 58 people in the control group would be below the average for the treatment group.

For a medium effect size (r=0.3), the difference is more substantial. For example, if there were 100 people in the control group:
- Only 27 of them (around a quarter) would score as well as an average person in the treatment group;
- Whereas 73 people (almost three quarters) in the control group would be below the average for the treatment group.

For a large effect size (r=0.5), the difference is very substantial. For example, if there were 100 people in the control group:
- Only 12 people in the control group (just over one tenth) would score as well as an average person in the treatment group
- Whereas 88 people (almost nine out of ten) in the control group would be below the average for the treatment group.

---

[52] See https://cep932.wikispaces.com/Effect+Size for an explanation of how Pearsons *r* is calculated manually.
[53] An updated version of the module was provided directly to the reports' data analyst (Elsbeth Asbeek-Brusse) by the module author, to address a minor programming issue.
[54] This comes from the widely cited book: *Cohen, J. (1988).* Statistical power analysis for the behavioral sciences. *NY: Academic Press*
[55] The proportions described are derived from Coe, 2002: table 1: 'interpretation of effect sizes' https://www.leeds.ac.uk/educol/documents/00002182.htm

7.3   Appendix 3: Executive Summary of Qualitative Study

How are English Language Teachers supported to develop professional knowledge and practice, within their schools?

Tom Power, Rama Mathew, Malcolm Griffiths (2017). English-in-Action, Dhaka, Bangladesh.

Introduction

This was a small-scale, qualitative study into the experiences of 'support in school for improving knowledge and practice of ELT' amongst teachers participating in a large-scale, quantitative, quasi-experimental (QE) study (EIA, 2017). The purpose of the study was to identify the nature of 'support in school' as experienced by teachers.

The study sought to explore aspects deliberately incorporated in the EIA School-Based Teacher Development (SBTD) programme (the *treatment* in the QE study) but which could also occur through other mechanisms (in *control* schools) such as:

- Teachers purposefully introducing communicative language teaching activities that were previously unfamiliar to themselves or their students.
- Teachers individually or collaboratively studying, planning, practicing or reflecting upon activities to improve their own English language proficiency, or that of their students
- The attitudes and actions of those in positions of authority in schools, such as head teachers and education officers.

The study addressed the following research questions:

1. In the setting of the school[1], how are English Language Teachers supported to develop their subject or pedagogic knowledge and their classroom practice?

2. How do contextual factors affect English Language Teachers' experiences of support in the setting of their schools?

3. Are there identifiable relationships between English Language Teachers' qualitative experiences of support in the setting of their schools, and the findings of quantitative studies of classroom practices or student learning outcomes?

[1] i.e. not at training courses or professional development events out-of-school

Methodology

The main research method was semi-structured interviews with teachers, informed by prior observation of their classroom practice. Semi-structured interviews were also carried out with head-teachers and education officers. Contextual information was gathered via questionnaire.

Eight schools, head teachers and education officers participated in the study, with sixteen English Language Teachers, distributed across conditions as shown in the table below.

	School performance in QE study	
	High Scoring	*Low Scoring*
QE Control	1 primary 1 secondary	1 primary 1 secondary
QE Treatment	1 primary; 1 secondary	1 primary; 1 secondary

There were two rounds of fieldwork (October 2015 and October 2016) during the study.

Findings

RQ1. There were marked differences between primary and secondary sectors. In all four primary schools, there was evidence of ongoing, active support. For example:

- In school 1 (high-scoring control) both teachers and head-teacher regularly met together to review and discuss classroom practices. The head teacher showed an awareness of communicative approaches to ELT ('*how to teach students English, using English conversation*') and actively encouraged greater use of English language. The Education Officer also promoted more student talk in English, as well as encouraging teachers to regularly sit together and discuss their teaching.
- School 2 (low-scoring control) began practicing 'Teacher Support Network' (TSN, a form of *Japanese Lesson Study*) during the study. Teachers met to plan

and review lessons which the head teacher observed, twice a month. Education Officers regularly visited school, observed lessons and promoted TSN activities.

- In school 3 (high-scoring treatment) the teacher had deliberately increased the use of spoken English over the year 'We think… they won't understand, but they understand'. Twice a month, teachers met together at lunchtime, to watch professional development videos and plan teaching. The Head Teacher observed and reviewed lessons with teachers. The Education Officer was actively supportive.
- In school 4 (low-scoring treatment) there was a strong collaborative culture, promoted by the head teacher. Teachers regularly met to watch professional development videos 'at tiffin time'. Some also studied these at home. One teacher referred daily to the teacher guide.

In three of the four secondary schools (both control schools and the low-performing treatment school) there was no indication of teachers' experiencing discernible support:

- School 5 (high-scoring control), no regular support was reported.
- School 6 (low-scoring control), the head teacher met with teachers but 'when they sit to discuss classes, it's just a formality'. Classroom practice is described as 'cramming, without understanding…'.
- In school 8 (low-scoring treatment), the head teacher says there is collaboration, but appeared dismissive. No regular support was reported by teachers.

In one secondary school (7, high-scoring treatment) teachers were discussing their teaching together, actively using professional development videos, audio resources and posters. The head teacher was actively promoting these activities.

RQ2. The most significant contextual factor was the support of the head teacher (and education officer) in promoting shifts in classroom practice and in fostering professional collaboration and support. In three of the treatment schools (3,4 and 7) professional development resources designed to facilitate support in school were widely used, whilst the fourth school (8) demonstrates the importance of school leadership to support this.

RQ3. Support in school does appear to be associated (tentatively) with higher scores in the QE study, whilst the absence of support appears associated with lower scores. In most higher-scoring schools (1, 3 and 7) there was evidence of a collaborative teacher development culture, promoted by strong school leadership; whilst there was only one high-scoring school (5), where this support appeared absent. Both low-scoring secondary schools (6 and 8) offered no discernible support. Whilst there was evidence of a collaborative teacher development culture being promoted by strong school leadership both low-scoring primary schools (2 and 4) this was only an established practice in one of the schools (4).

institutionalisation or contamination?
Almost half of the case study schools provided evidence of ways in which classroom or teacher development activities promoted through the treatment were being actively spread beyond the treatment schools, by education officers, head teachers and teachers. Whilst this represents a 'success story' for institutionalization, it is also 'contamination' in the sense of a QE study:

- In school 1 (primary control) the teacher reported 'her daughter's school follows EIA (treatment) techniques and she has adapted many of those in her own class'. The Head Teacher (HT) was aware of EIA (treatment) and wanted training in 'how to teach students English, using English conversation'; the HT encouraged teachers to speak English and allow students to practice English. The Education Officer (EO) liked the use of teacher peer-support, supplementary classroom materials and pair and group work in EIA (treatment) and was actively promoting these throughout all their schools.
- In school 2 (primary control) both teachers reported taking part in the new Subject-Based Training and sharing what they had learned across the school. This government programme has aspects of the treatment institutionalized within it.
- In school 4 (primary treatment), the EO said they actively promoted EIA (treatment) approaches to all schools.

Conclusions

In primary schools, teachers experienced support through a variety of mechanisms. In secondary schools, such support was rare, only evidenced in a treatment school (7). Activities and resources provided through the treatment (EIA) were used regularly in schools, both individually and collaboratively. The role of head-teachers (and to some extent, education officers) was important both in terms of fostering a collaborative, developmental environment, and in focussing upon learning and teaching. Schools where such support was established tentatively seem to be associated with higher scores in the QE study.

7.4 Appendix 4: GESE and Common European Framework of Reference (CEFR).

What are CEFR and GESE?

The *Common European Framework of Reference* (CEFR) *for languages* (Council of Europe, 2011) has been under development by the European Union 'Language Policy Unit' since the early 1970s. It is intended to provide a framework of reference that can be applied to the learning, teaching and assessment of any European language (it is not specific to English).

CEFR identifies six levels of language proficiency, with level descriptors being developed pragmatically, from descriptions that have been found to be transparent, relevant and useful to teachers and learners from a broad range of sectors and linguistic backgrounds, including native and non-native language learners (p6). The levels are

- Level A: Basic
- Level B: Independent[56]
- Level C: Proficient

Each level is divided into two numbered sub-levels, with level one (e.g. A1) being the lower and level two (e.g. A2) being the higher level. Summary descriptions are provided in 'holistic paragraphs' for representation with non specialist users (p24). CEFR *does not* provide or promote any one internationally recognised method of language proficiency assessment, even in English. Rather, it provides a common framework against which various teaching and assessment schemes may be considered, either within or between different modern European languages.

The *Graded Examinations in Spoken English* (GESE) are a specific set of English language examinations designed by Trinity College London (2014), to provide qualifications in English for Speakers of Other Languages (ESOL), mapped onto the six CEFR levels from A1 to C2. GESE is designed for use worldwide and is internationally recognised as providing '…*a reliable and valid scheme of assessment through which learners and teachers can measure progress and development*…' (p5).

In GESE, progress is marked by:

- Increasing evidence of the use of communicative skills
- Greater linguistic demands
- Increasing learner independence

GESE has four broad levels (Initial, Elementary, Intermediate and Advanced) each divided into three finer grained levels, to give a twelve-point scale (p. 6). One benefit of the GESE system is that it is intended to provide a motivational tool, encouraging learner development by acknowledging small steps, with opportunity to mark progress even at very early language development stages. To this end, GESE Grade 1 acknowledges achievement that precedes the competencies described in CEFR

---

[56] Sometimes referred to as 'Intermediate' rather than 'Independent' user level.

A1. (Although not widely publicised, advanced assessors are also able to assess each of the twelve levels to three further sub-levels, allowing the potential for a thirty-six-point assessment scale.)

Why did EIA choose GESE to assess English Language Proficiency?

GESE was attractive as it was internationally recognised, with well-established procedures for assessor verification and assessment validity, as well as having a large network of trained and moderated assessors, many of whom were experienced in working with early-stage language learners in South Asia (and some based in India). The GESE progress criteria were also well aligned with the pedagogic intent of the EIA programme and Government of Bangladesh policy for English language learning.

The assessment itself provided for a much higher degree of discrimination than the broad CEFR levels. The finer granularity was more likely to detect the relatively small improvements in language proficiency anticipated as a result of students participating in new classroom activities over perhaps a six-month period, whilst their teachers took part in a school-based professional development programme for one year.

It was also judged to be important, particularly for primary students, that the assessment method should cover *pre-CEFR A1* language learning, as it was suspected that for many students, CEFR A1 represented a medium-term development goal that was unlikely to be achieved during their teachers' participation in a 1 year CPD programme.

How are GESE levels assessed

The standard GESE examination protocols have been slightly modified to suit the EIA context: candidates don't 'enter an exam' to be assessed for one particular level and at higher GESE levels the candidate led topic discussion and interactive tasks are omitted. The GESE assessment for EIA takes the form of a 'diagnostic interview', which begins with the kinds of language function and lexis described in GESE level 1, such as the exchange of greetings with the examiner, showing understanding of simple instructions (such as sit down) with appropriate actions, and giving short answers to simple questions (e.g. 'what's your name'). If the candidate is comfortable and competent in these exchanges, the assessor gradually increases the communicative and linguistic demand of the conversation, gently working stepwise up through GESE levels, until the candidate is no longer able to continue extending the conversation. At this point, the interview is drawn to a close and the assessor notes the highest level at which the candidate was comfortable and confident to communicate.

How do GESE levels map onto CEFR?

There has been extensive work to map and verify the GESE levels against CEFR (Trinity College London, 2007). The mapping is provided in the table below.

Common European Framework of Reference (CEFR)	Graded Examinations in Spoken English (GESE)	Integrated Skills in English (ISE) exams	Spoken English for Work (SEW) exams
–	Grade 1	–	–
A1	Grade 2	–	–
A2	Grade 3	ISE 0	–
A2	Grade 4	ISE 0	–
B1	Grade 5	ISE I	SEW B1
B1	Grade 6	ISE I	SEW B1
B2	Grade 7	ISE II	SEW B2
B2	Grade 8	ISE II	
B2	Grade 9	ISE II	SEW B2+
C1	Grade 10	ISE III	SEW C1
C1	Grade 11	ISE III	SEW C1
C2	Grade 12	ISE IV	–

After Trinity College London (2014):60

References

Council of Europe (2011). COMMON EUROPEAN FRAMEWORK OF REFERENCE FOR LANGUAGES: LEARNING, TEACHING, ASSESSMENT. Technical report, Language Policy Unit, Council of Europe, Strasbourg. http://www.coe.int/t/dg4/linguistic/Source/Framework_EN.pdf

Trinity College London (2007). Relating the Trinity College London GESE and ISE examinations to the Common European Framework of Reference – project summary. Technical report, Trinity College London, London. www.trinitycollege.com/resource/?id=2261

Trinity College London (2014). Exam Information: Graded Examinations in Spoken English (GESE). Technical Report Sixth Impression, Trinity College London, London, UK. http://www.trinitycollege.com/resource/?id=5755

ENGLISH in ACTION

f /EnglishInActionBangladesh
twitter.com/EIABD
youtube com/eiabd
flickr.com/photos/eiabd
info@eiabd.com
www.eiabd.com

"English in Action (EIA) is a UK Government funded programme implemented by the Government of Bangladesh and managed by Cambridge Education, a member of Mott MacDonald."

For more information,
camb-ed.com