Ministry of Housing, Communities & Local Government

Measuring the impact of Community-Based English Language Provision

Findings from a Randomised Controlled Trial

1 Abstract

This report presents findings from a randomised controlled trial (RCT) of a Community-Based English Language (CBEL) intervention aimed at people with very low levels of functional English proficiency. The intervention consisted of 66 hours of guided learning and support delivered through 22 classes and 11 club sessions over an 11-week period. The programme sought to test whether the intervention was effective in supporting individuals from communities with very low levels of functional English to improve their ability to communicate in English, and integrate into their wider community and society. This trial sought to test the effect of the intervention on English language proficiency and social integration. Integration was measured in terms of social interactions and mixing; participation in everyday activities; confidence in engaging with public services; local and national belonging; trust in others and attitudes to community integration.

A two-armed randomised controlled trial was deployed, with a waiting list design for the control group. All trial participants received the intervention, with those assigned to the treatment group receiving the intervention over an 11-week period between April and July 2016, and those assigned to the control group receiving the intervention from September 2016. In total, 527 participants were recruited to the trial.

Measures of English language proficiency and social integration were taken at the beginning and end of the intervention period (prior to the control group having attended any class sessions). The intervention was delivered by a consortium of partners led by Manchester Talk English (part of Manchester City Council) at 22 community venues operating across five local authorities.

The programme was designed and implemented by the Ministry of Housing, Communities and Local Government (MHCLG), with expert advice from the Behavioural Insights Team (BIT) and the Cross-Government Trials Advice Panel. The trial was carried out by the Learning and Work Institute, working in partnership with BMG Research.

The trial found a strong and sizable difference in overall English proficiency and the amount of progress made between the treatment and control groups across all the English proficiency domains (speaking and listening, reading and writing).

Significant differences were also observed between the treatment and control groups in relation to changes in the number of social interactions and friendships being formed, levels of trust in other people, confidence in talking to health professionals and visiting shops independently. These findings suggest that the CBEL intervention increased learners' social mixing and, to a more limited extent, their participation in wider society. Overall, findings indicate that the provision of Community-Based English Language support can improve English proficiency and also promote social integration.

Contents

Acknowledgements

Authors

Ash Patel is Head of Research (Inclusion) at the Learning and Work Institute (L&W). Ash led the quantitative strand of the evaluation, overseeing the learner survey and impact assessment.

Carmen Hoya was formerly a Researcher at L&W. Carmen managed the process evaluation.

Paul Bivand is Associate Director of Statistics and Analysis at L&W, with 30 years of experience in employment and skills policy research. Paul leads on the analysis of existing survey resources, and where these are inadequate; he manages specialised surveys commissioned by clients.

Alex McCallum was formerly a Labour Market Researcher and Economist at L&W. Alex supported both strands of the research.

Alex Stevenson is Head of English, Maths & ESOL at L&W, leading on the development of a Citizens' Curriculum approach in adult learning.

Tony Wilson is Director of Policy and Research at L&W. Tony directed the evaluation and had oversight of both strands and all findings.

Glossary

English for Speakers of other languages (ESOL)

Pre-entry level English proficiency

Pre-entry level denotes a very low level of English proficiency. An individual with pre-entry English may be able to answer questions on basic personal information and follow basic instructions but would have very limited (if any) letter and word recognition.

Entry level 1 English proficiency

Entry level 1 denotes a very basic level of English proficiency. An individual at this level may be able to ask and respond to personal information questions (in more than one word answers); give short accounts of activities; and make simple statements of fact. This level equates to standards of literacy and language expected of native speakers aged 5 to 7.

Entry level 2 English proficiency

Entry level 2 denotes a basic level of English proficiency. An individual at this level may be able to answer questions about their daily routine; give short accounts of previous experiences; and ask similar questions with the correct verbs and tense. This level equates to standards of literacy and language expected of native speakers aged 7 to 9.

English proficiency assessment

A proficiency assessment is a test to measure a learner's level of language and ability to use English.

Randomised Controlled Trial

Randomised Controlled Trials (RCTs) are a research method used to establish impact. They involve a control group who does not receive an intervention and effects are compared with those who do. Participants are randomly allocated to each group.

Randomisation

The process by which participants in an RCT are allocated to the treatment or control group (to receive or not receive the intervention). Randomisation is a critical element of an RCT as it ensures there are no systematic selection biases between participants allocated to either group. This helps to remove bias or interference caused by other factors. The result of randomisation will be that the two groups share, on average, very similar characteristics.

Blinding

Blinding refers to concealing the allocation of participants to the treatment or control group. It is considered an optimal approach in RCTs as it minimises the likelihood of participants being treated or assessed differently based on their group allocation.

Trial participant   An individual eligible to enter the RCT and has provided informed consent to do so.

Treatment group   The group that receives the intervention within an RCT design following randomisation.

Control group   The group that does not receive the intervention within an RCT design following randomisation. They are monitored alongside the group receiving the intervention, and their results are compared to their treatment counterparts to understand what impact the intervention has had, compared to receiving no intervention. Any changes or effects detected within the control group over the course of the RCT can be interpreted as what would have happened anyway.

1 Executive Summary

1.1 Introduction

This report presents findings from a randomised controlled trial (RCT) of a Community-Based English Language (CBEL) intervention aimed at people with very low levels of functional English proficiency. The intervention consisted of 66 hours of guided learning and support delivered through 22 classes and 11 club sessions over an 11 week period. The programme sought to test whether the intervention was effective in supporting individuals from communities with very low levels of functional English to improve their English proficiency, and encouraging them to integrate into their wider community and society. For the purposes of this trial English language proficiency is considered to be the primary outcome measure. Social integration outcomes are considered to be secondary measures.

1.2 Methodology

A randomised controlled trial design was implemented using a waiting list design for the control group. Participants assigned to the treatment group received the intervention over an 11-week period between April – July 2016. The control group received the intervention from September 2016, once the trial had concluded. Measures of English proficiency and social integration were taken at the beginning and end of the intervention period. The intervention was delivered by a consortium of partners led by Manchester Talk English (the provider) through 22 community centres operating across five local authorities.

1.2.1 Trial participants

The trial targeted, and was successful, in recruiting people from communities with very low levels of functional English. By delivering the intervention in specific geographic locations, the trial sought to specifically engage women from the Pakistani, Bangladeshi and Somali communities (as these groups have the lowest self-reported levels of fluency in English)[1], though anyone satisfying the eligibility criteria was able to join. Trial participants were engaged by the provider, and recruitment was carried out through a series of registration events arranged prior to the start of the course.

1.2.2 Intervention

The intervention was designed to improve functional English proficiency. Classes were held in community settings (e.g. community centres, libraries, family centres, etc.) and delivered outside of a usual adult education location. The CBEL intervention is similar to existing Community-Based ESOL provision, although standard provision is usually less

---

[1] DCLG. (2011). *Citizenship Survey 2010-11*. Unpublished analysis of rates of (self-reported) below average or poor spoken English by ethnic group and country of origin. Data available at: https://discover.ukdataservice.ac.uk/catalogue/?sn=7111&type=Data%20catalogue. (Accessed: February 03, 2012); ONS. (2011). Unpublished analysis of ethnic group by proficiency in English by sex by age (Table CT0558). Data available at: http://www.ons.gov.uk/ons/about-ons/business-transparency/freedom-of-information/what-can-i-request/published-ad-hoc-data/census/ethnicity/ct0558-2011-census.zip. (Accessed: January 21, 2016).

structured, *usually* of lower intensity and not as specifically targeted. Some standard provision also relies more on unqualified volunteer teachers rather than paid qualified staff.

1.2.3 Trial implementation

Overall 527 participants were recruited for the trial of which 249 were randomly assigned to the treatment group and 278 to the control. Retention within the trial was high and in line with expectations (as outlined in the trial protocol)[2], with 61.9 per cent of all participants providing data at all stages of the study. The implementation of the intervention was closely monitored by the provider and MHCLG. An accompanying process evaluation was also conducted to review the process of delivering the trial. This found that overall the trial was implemented effectively, with fidelity to the experimental design and to the design of the course itself.

While outcome measures were consistently collected for most learners, reading and writing measures were not fully captured for all learners at one centre. As a result, analysis of the reading and writing proficiency excluded all learners from this centre. The trial sample was sufficiently large that this had minimal effect overall.

1.2.4 Condition balance

Checks of the balance in characteristics of the treatment and control groups at baseline and follow-up stage indicated that both groups appeared well balanced with regards to their socio-demographic profile throughout the trial period. Likewise, exploration of baseline scores found no statistically significant variation in speaking and listening and writing scores between the two groups. There was however a significant difference between groups in reading scores at baseline. The reasons for this are unclear. The difference, however, does not appear to be related to the initial randomisation of trial participants. While this baseline difference should be noted, analysis of the impact of the intervention controlled for baseline variation, thereby removing any potential distortion.

1.3 Results

Findings suggest strong and highly significant impacts across all of the English proficiency domains following receipt of the intervention. Analysis of the social integration outcomes indicated that the intervention led to improvements in social interaction and bond forming. There was also evidence of increased confidence in independently going shopping and engaging with health professionals but not for some other activities such as talking to a teacher or the police.

Overall, these findings indicate that Community-Based English Language courses increase English language ability and encourage wider social integration.

1.3.1 English proficiency score

The trial employed two measures of English proficiency on which the treatment and control groups were compared (across each of the speaking and listening, reading and writing domains).

---

[2] The trial protocol can be found in Annex A of the report.

Mean change in score measures the change between pre and post assessment scores and indicates the rate of improvement in English proficiency over the course of the trial. Follow-up score measures proficiency at the end of the trial, indicating the overall ('absolute') level of English proficiency attained by each group.

Attending the intervention course significantly improved English proficiency on both outcome measures.

1.3.2 Mean change in English proficiency score

Improvement in English language proficiency among the treatment group exceeded that of the control group. For speaking and listening scores, the level of improvement within the treatment group was double that achieved by the control group (respectively, the mean change was 1.39 points compared to 0.71 (t(324)=3.653, p<.001). Similar degrees of change were observed in both the writing and reading scores, though the change in the latter should be viewed with caution considering the significantly higher reading scores among the treatment group at baseline.

1.3.3 Follow-up English proficiency score

Comparing follow-up English proficiency scores showed sizeable differences between the treatment and control groups with higher mean follow-up assessment scores across all outcome measures. For speaking and listening, the treatment group scored on average almost a full one point, or the equivalent of a third of an English proficiency level, higher than the control group (respectively, the mean follow up score was 5.91 compared to 4.92 t(351)=4.928, p<.001).

Improvements in both the writing and reading assessments appeared similarly large in scale. The difference in follow-up reading scores should however be viewed with caution given the significantly higher reading scores among the treatment group at baseline.

1.3.4 Key predictors of progess in English proficiency

A set of linear multilevel models were fitted to data, to identify factors associated with the change in English proficiency score. Having a university education, a child over the age of 5, English proficiency score at baseline and, importantly, attending the CBEL course were all key predictors of change in the speaking and listening score.

English proficiency score at baseline was a strong predictor of change in score, with higher baseline scores associated with smaller gains in proficiency (across all measures).

Controlling for all other variables in the model, attending the CBEL course increased the change in the speaking and listening score of the treatment group by an additional 0.92 points (t=3.70, p<.001) relative to the control group (the change in score was nearly double for the treatment group when compared to the control at 2.06 and 1.14 respectively). Put another way, this is the equivalent of a treatment group learner who comes in at the lowest Entry Level 1 score (a score of 4 on the scale) progressing over two-thirds of the way to the next proficiency level (Entry Level 2) within the duration of the course.

Models were also fitted to identify predictors of change in reading and writing proficiency and showed similar predictors on both measures including trial condition.

1.3.5 Key predictors of overall English proficiency

Statistical models were also fitted to data to identify factors associated with follow-up English proficiency score. Again, university education, having a child over the age of 5, baseline English score and attending the course were all key predictors of follow-up speaking and listening score.

Baseline English proficiency was also found to be a strong predictor of the follow-up score achieved (higher scores at baseline were predictive of higher scores at follow-up).

Controlling for all other variables in the model, attending the CBEL course increased follow-up scores of the treatment group by 1.2 points (t=3.65, p<.001) relative to the control group. This is equivalent to a 13 percent higher follow-up score among the treatment group compared the control group.

Models were also fitted to identify predictors of follow-up reading and writing proficiency and showed similar predictors.

1.3.6 Social integration

A two wave survey was carried out to explore the effect of the intervention on social integration outcomes. The survey asked about participants' social interactions; everyday activity; confidence in engaging with public services; independence; trust in others; attitudes to community integration and belonging; and interest in future training opportunities.

Significant differences were observed between the treatment and control groups in relation to changes in the number of social interactions and friendships being formed, levels of trust in other people, confidence in talking to health professionals and visiting shops independently.

These findings suggest that the CBEL intervention increased learners' interactions using English outside of their class, increased their social mixing and to a more limited extent their participation in wider society. Overall, findings suggest that the provision of Community-Based English Language support can promote social integration.

No differences were found between the treatment and control groups' attitudes to community integration, their sense of independence or belonging (locally or nationally) or their interest in attending future training. However, this may in part be due to the very high rates of belonging, personal agency and interest in training reported by both groups.

These results are promising as it may reasonably be expected that levels of confidence will increase further, as learners engage in more activities for themselves and have greater social interaction with people from different backgrounds.

2 Introduction

This report presents findings from a randomised controlled trial (RCT)[3] of a Community-Based English Language (CBEL) intervention. The intervention consisted of 66 hours of guided learning and support delivered through 22 classes and 11 club sessions over an 11 week period. The course was specifically targeted at people with very low levels of functional English proficiency.

The programme was designed and overseen by the Ministry of Housing, Communities and Local Government (MHCLG) who were also responsible for the design of the trial with input from the Behavioural Insights Team. The Learning and Work Institute, in partnership with BMG Research, were commissioned to implement RCT procedures, including the collection of measures and analysis. Oversight for the evaluation was provided by advisors from the Cross-Government Trial Advice Panel including Government and academic experts in the field of impact evaluation.[4]

Overall, within government it is the Department for Education (DfE) that is responsible for funding English for Speakers of Other Languages (ESOL) provision through the Adult Education Budget. Full-funding for this provision is prioritised for the unemployed on benefits, whose poor command of English is a barrier to getting a job. In 2014/15, DfE invested an estimated £104 million on fully and part-funded ESOL courses, supporting 131,000 adult learners. MHCLG is responsible for policy on integration. Supporting people to learn English has been a core part of MHCLG's approach which is set out in *Creating the Conditions for Integration* published in 2012.[5]

Evidence shows that poor English proficiency is a barrier to both economic and social integration. The 2011 Census found that over 760,000 adults born outside the UK and living in England and Wales cannot speak English well or at all.[6]

From 2013/14 to 2015/16, MHCLG funded an £8m Community-Based English Language programme, supporting six projects to deliver English courses to adults with the lowest levels of English. The projects operated in the English language priority areas: broadly in East and North London; East Birmingham; Manchester; towns along the M62 in Yorkshire and Cheshire; Slough; Luton and Bristol. The projects were selected for their innovative teaching and engagement models, which delivered training in community-settings or on-line often using volunteers. Together the projects reached over 39,000 adults - around 80 per cent of whom were women, with over half from Pakistani, Bangladeshi, and Somalian ethnic groups.

---

[3] For reasons of practicality and to ensure the intervention was viable, randomisation was clustered by groups of known learners registered on the course (discussed below).
[4] For more information about the Cross-Government Trial Advice Panel, see
https://www.gov.uk/government/publications/cross-government-trial-advice-panel-role-and-membership
[5] For more information about the approach published under the 2010 to 2015 Conservative and Liberal Democrat coalition government, see:
https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/7504/2092103.pdf
[6] ONS (2011) English Language Proficiency by Age by Sex by Country of Birth by Year of Arrival in the UK (Table BD0059). Data available from: http://www.ons.gov.uk/ons/rel/census/2011-census-analysis/social-and-economic-characteristics-by-length-of-residence-of-migrant-populations-in-england-and-wales/rft7---bd0059.xls

To strengthen the evidence base MHCLG commissioned this Randomised Controlled Trial to test the impact of a Community-Based English Language intervention. Projects were invited to submit proposals for delivering an intervention on an RCT basis, and Manchester Talk English's proposal was selected.

2.1 Evaluation aims

This trial tested whether an intensive English language programme delivered within a community setting (described in more detail below) could improve both the English language proficiency and the social integration outcomes of participants. A RCT is commonly considered the strongest form of impact evaluation. It is accompanied by a qualitative process evaluation examining the implementation of the intervention and the RCT.[7]

The overall aims of the programme were to:

1. increase English language proficiency of programme participants;

2. increase social integration amongst programme participants.

The trial therefore sought to address the following research questions:

1. Do individuals who attend an 11-week Community-Based English class have significantly better levels of English Proficiency than individuals who do not?

2. Do individuals who attend an 11-week Community-Based English class have significantly better improvement rates in their English Proficiency than individuals who do not?

3. Do individuals who attend an 11-week Community-Based English class show different levels of integration (shown through social mixing, participation in wider society, etc.), to those who do not and how has this changed over time?

As the intervention was designed to specifically improve functional English language proficiency, the primary outcome measure for the trial was the impact on English language proficiency. Social integration measures were considered secondary outcomes.

---

[7] https://www.gov.uk/government/publications/community-based-english-language-programme-a-randomised-controlled-trial--2

3 Methods

The research strategy was developed by MHCLG with input from a range of specialists and programme and evaluation partners. In collaboration with the Behavioural Insights Team, MHCLG developed an RCT protocol (see Annex A) which underpinned the design and implementation of the trial.

A description of how the trial was implemented is provided below.

3.1 Trial design

The programme employed a two-arm randomised controlled trial, using a waiting list design for the control group. Accordingly, all trial participants received the intervention, with those assigned to the treatment group receiving the intervention over an 11-week period between April – July 2016. The control group received a similar intervention beginning September 2016, around six to eight weeks after the treatment group had completed the course and follow-up measures had been obtained from both groups.

Manchester Talk English were commissioned by MHCLG to develop and deliver a Community-Based English Language intervention, derived from their existing Talk English programme, which could be delivered under randomised controlled trial conditions. They designed the overall programme, developed the intervention manual and teaching materials, recruited and trained teachers, volunteers and support staff, sourced community partners and venues to deliver the intervention and oversaw the recruitment and registration of CBEL trial participants. They were also responsible for ensuring consistency in delivery of the intervention across all areas.

The trial was run within five local authority areas: Bradford; Kirklees; Manchester; Oldham; and Rochdale. It was delivered at 22 community centres, with one centre running two classes. It was originally anticipated that the trial would be run with 25 classes, but two classes were not able to recruit sufficient participants (a minimum of 18 learners per centre was required prior to randomisation for the intervention to be viable in those locations).

Figure 1 provides a high-level overview of the structure of the trial. It shows how the trial took place in multiple delivery centres and randomisation took place at each centre to create a treatment (T) and control group (C). The control group were placed on a waiting list, and offered the intervention after the treatment groups had completed.

Figure 1: Structure of the CBEL trial
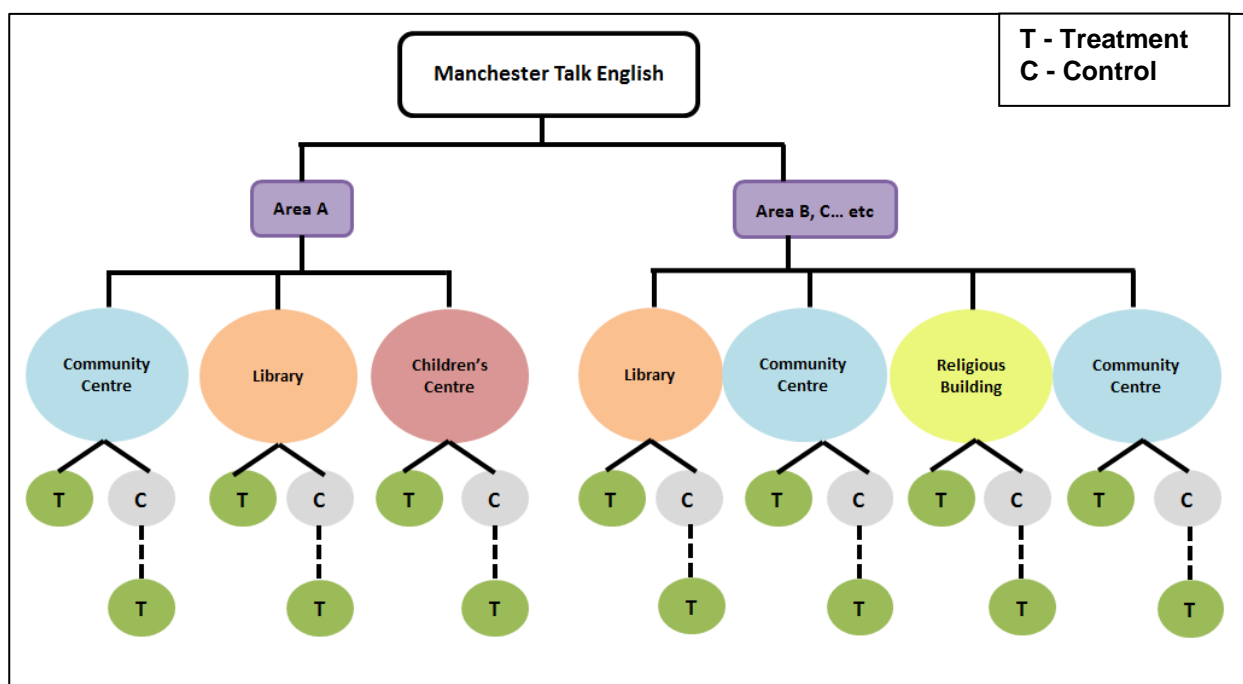
3.2 Participants

Trial participants were engaged by the provider through a variety of activities (more detail about the engagement strategy can be found in the accompanying RCT process evaluation report).[8] Assessment of trial participants' eligibility and recruitment to the trial (including collection of informed consent) was carried out through a series of registration events which occurred between two and five weeks prior to the commencement of the course.

Inclusion criteria were specified by MHCLG. Trial participants were expected to:

1. have low levels of English language proficiency (equivalent to Pre-entry level or Entry level 1 on any **one or more** of the assessment areas (speaking and listening, reading or writing) as defined by the ESOL classification criteria)[9];

2. be resident in one of the 5 local authority areas the trial was operating in;

3. not be eligible for other English language provision (for example, through entitlements through their employers or through eligibility for Job Seekers Allowance);

4. be aged 19 and above;

5. have been resident in the country for more than 12 months;

6. not have received formal support from Talk English in the past (e.g. a Talk English course);

7. consent to participate in the research.

---

[8] Process evaluation of community-based English language provision
[9] For more information about the core ESOL curriculum visit: http://www.excellencegateway.org.uk/content/etf2385

Recruitment guidelines additionally encouraged the recruitment of women from Pakistani, Bangladeshi and Somali communities (as these groups have the lowest self-reported levels of fluency in English).[10] In order to maximise the number of participants from these backgrounds, the trial was deliberately run in local areas with high concentrations of these population groups.

Due to initial difficulties in recruiting sufficient numbers of trial participants, some eligibility criteria were relaxed. After careful consideration providers were permitted to invite individuals who had had previous interactions with the provider, but had not received any form of taught support, to register for the trial. This included individuals who had previously enquired about English language courses and/or had attended a taster session or drop in 'café' session.

Eligibility was assessed by the programme provider at registration events using the provider's standard English language proficiency assessment screening tools, and self-reported information from the potential participants.

3.3 Intervention

The CBEL intervention was an 11-week course designed to improve functional English proficiency. The intervention was delivered in community settings rather than adult education facilities (such as further education colleges) in order to appeal to individuals who would not access support in more formal settings. Courses operated out of a range of community venues, including women's centres, community centres, churches, and family/children's centres.

Learners were expected to attend three sessions per week – two sessions of the Talk English Together classes and one session of the Talk English Together club (both elements were compulsory). All sessions lasted around two hours, comprising 66 guided learning hours in total. Classes ran twice a week (accounting for 44 guided learning hours). Classes were taught by a qualified ESOL teacher, with additional support provided by up to two regular volunteers. They were classroom based and focussed on developing the English language skills of learners, increasing their confidence in using English language and their participation in the community.

Clubs ran once a week (accounting for 22 guided learning hours). Clubs were supported by two volunteer Talk English Friends and supervised by the qualified ESOL teacher. They provided learners with the opportunity to put their language skills into practice and to gain experience of their local community (through visits to and talks from local service providers).

The Talk English Together course was modular in design. Content for each class was prescribed by a course manual outlining the objectives of the class, issues to cover and

---

[10] DCLG. (2011). Citizenship Survey 2010-11. Unpublished analysis of rates of (self-reported) below average or poor spoken English by ethnic group and country of origin. Data available at: https://discover.ukdataservice.ac.uk/catalogue/?sn=7111&type=Data%20catalogue. (Accessed: February 03, 2012); ONS. (2011). Ethnic group by proficiency in English by sex by age (Table CT0558). Data available at: http://www.ons.gov.uk/ons/about-ons/business-transparency/freedom-of-information/what-can-i-request/published-ad-hoc-data/census/ethnicity/ct0558-2011-census.zip. (Accessed: January 21, 2016).

providing class materials for individual sessions (a copy of the manual can be found in Annex D).

Nine topics (and associated vocabulary) were covered over the 11 week course. These included friends and family, health, food and shopping, community participation and democracy. Modules were delivered in a standardised order and with a standardised amount of time devoted to each module (no flexibility was permitted in the order or time spent covering a specific topic).

The CBEL intervention trialled contrasts with other Community-Based ESOL provision, which is sometimes less structured and standardised, of lower intensity (e.g. one session per week), not as specifically targeted to particular learner groups, and can have a greater reliance on volunteer teachers.

It was anticipated that a maximum of 12 individuals would attend any one class, though as a result of randomising by cluster, it was not always possible to ensure an equal split of participants across classes/centres. As a consequence, the number of individuals assigned to any one centre ranged from a minimum of eight to a maximum of 15.

The start of the intervention was staggered, with centres beginning classes between 25th April and 9th May 2016. This timetable was adopted to ensure that the trial did not run over the school holidays (specifically the summer break) which it was anticipated would substantially decrease the likelihood of participants with school aged children attending. Control group participants commenced their course in September 2016.

3.4 Outcome measures

Outcome measures were obtained by a series of English language proficiency tests which were developed by the English Speaking Board (ESB) and administered by ESB assessors and CBEL tutors, and a paper based survey administered by BMG research (on behalf of L&W). These measures were obtained twice for both groups. Pre-intervention measures were collected as close to the commencement of the trial as possible (in most instances during the first week of treatment group classes). 'Post' measures were collected as close to the conclusion of the learning as possible (in most instances during the final week of treatment group classes).

The initial trial protocol (as outlined in Annex A) had also factored in taking interim English proficiency assessments six weeks through the course from a small sub-sample of learners. These were not intended to contribute to the final outcomes, and were only to provide a descriptive indication of progress. However, logistical difficulties and time pressures prevented these interim measures from being taken.

The primary outcome measure, English language proficiency, was assessed by bespoke tests (copies of which are available on request). The tests provide a ten-point scale on which English proficiency in speaking and listening, reading and writing can be measured, where the lowest score is 0 and the highest 9 (further detail about the interpretation of scores can be found in Table 2, below). The scale and the tests were developed by ESB

for this project based on their expert knowledge of assessment methods and were consistent with the ESOL core curriculum.[11]

Social integration outcomes were analysed once the primary outcome measures had been completed by learners. These were obtained by a survey which comprised questions based on the following key themes:

1. Social interaction

2. Everyday activity

3. Confidence engaging with public services

4. Independence

5. Trust in others

6. Community integration and belonging

7. Interest in training and employment

The survey was developed by MHCLG with input from the Behavioural Insights Team and the Cross-Government Trials Advice Panel. Some of the questions were adapted from existing surveys (including the Citizenship Survey and European Social Survey) while others were created specifically for this research.

The survey was refined, cognitively tested and piloted by BMG Research. Twenty cognitive interviews were undertaken with individuals drawn from similar populations to CBEL eligible learners. Interviews were conducted in Urdu, Bengali, Punjabi, Somali and Arabic, testing comprehension of the questions, the appropriateness of the scales employed, and consistency in the understanding of the response options available. The revised survey was then piloted with a further 20 individuals drawn from similar populations to the target group for the trial.

The survey was translated into the five core languages (Urdu, Bengali, Punjabi, Somali and Arabic) to capture the main language groups spoken by participants. Only individuals who spoke one of these five languages were asked to complete the survey. One question was incorrectly translated on the Punjabi version, meaning that the data for this question was not comparable across language groups.[12] Respondents affected by this error were identified and excluded from the analysis of this specific question.

Further detail about the development of the survey can be found in Annex B, and copies of both the baseline and post-intervention follow-up surveys can be found in Annex C.

The trial protocol outlined plans to include questions on voting behaviour and to capture further data on learners' behavioural motivations through a sign-up sheet where learners

---

[11] For more information about the core ESOL curriculum visit: http://www.excellencegateway.org.uk/content/etf2385

[12] The Punjabi version asked '*Apart from your English class, how many people did you speak to last week from a different country or religion to you in English?*'. The reference to 'in English' was an error and was not present in the other language versions. This error did not become apparent until after the baseline survey had been delivered and completed by 51 respondents. For a more detailed analysis of this issue, see the survey development report contained in Annex B.

could indicate which local services they were interested in engaging with after the course. The voting questions were excluded from the survey due to concerns that the majority of learners would not be eligible to vote. Following further discussion with partners, the sign-up sheet was also not taken forward, as partners were concerned about learners misunderstanding the hypothetical nature of the sign-up sheet, and creating false expectations that such services would be immediately available to them.

3.5 Procedure

The provider, Manchester Talk English, was responsible for the engagement and recruitment of trial participants as well as the delivery of the intervention. In order to fulfil this, a consortium of partners was established with representation across the five participating local authorities. Figure 2 outlines the structure of the Talk English consortium, and the lines of responsibility.

Figure 2: Structure of the Talk English consortium
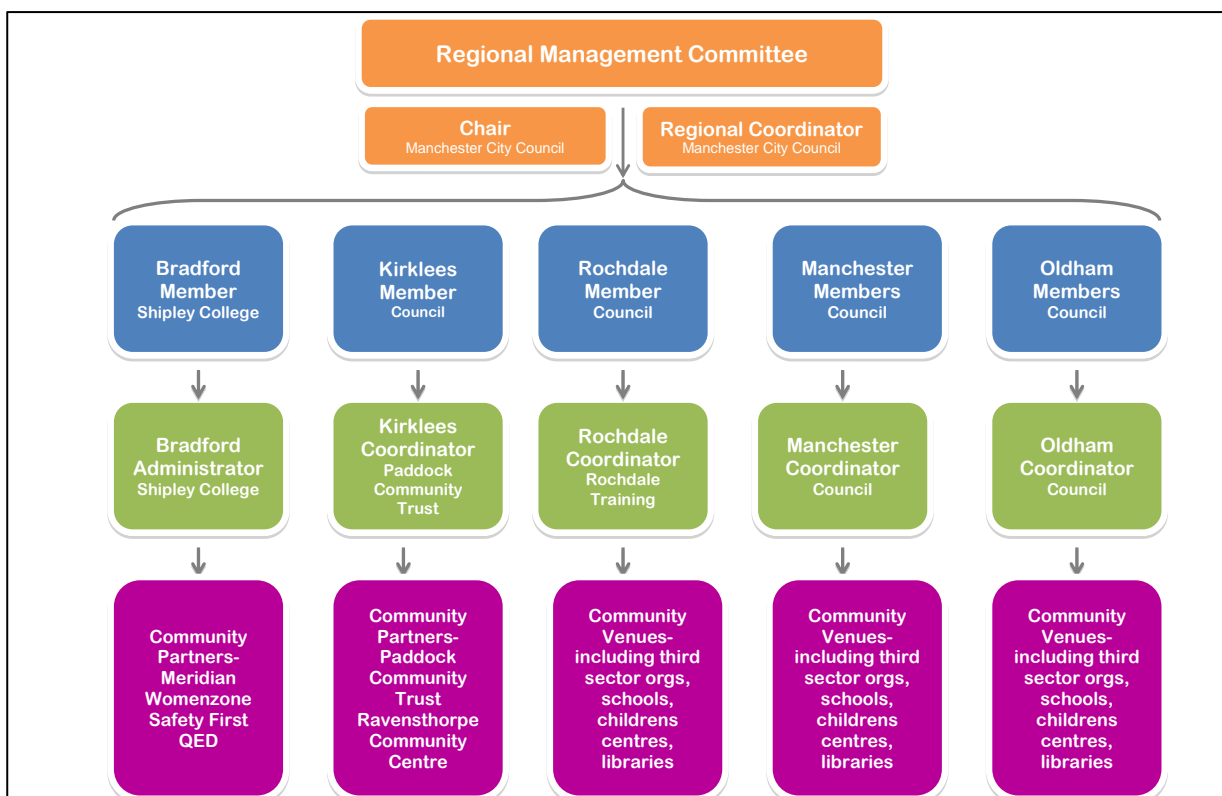
Recruitment of participants was carried out locally, by area co-ordinators and administrators who were also responsible for sourcing community venues, managing the registration events and ensuring that the course was delivered in line with the Talk English Together manual.

Registration events were held in the intervention delivery centres throughout March and April 2016. If individuals interested in attending the Talk English Together course met the eligibility criteria, a registration form was completed by the provider on their behalf. Each registration form was unique, and was pre-printed and supplied by BMG Research with a designated unique learner identification reference.

As part of the eligibility criteria, interested individuals had to agree to entering into the CBEL trial, participating in its research activities, and sharing their data with the evaluation partners and MHCLG. In order to gain informed consent, a form explaining the nature of their involvement and data sharing implications was provided and acknowledgment of their understanding was confirmed by their signature. The consent form was designed to be easy to read in plain English, and was translated into Arabic, Bengali, Punjabi, Somali, and Urdu. Further, for those unable to read the forms, audio recordings explaining the content of the forms were provided in the five languages and a telephone interpretation service was made available to centre staff (for assistance with additional languages). Individuals who did not speak one of the five languages, were invited to take the form home and have it translated by a friend or family member, if there was no such facility available at the registration event itself.

The trial set out to recruit 600 participants across 25 centres (equivalent to 24 participants per site). In total, 527 participants were recruited to the trial. Once registered, participants were randomised by L&W into either treatment or control groups (discussed below).

3.6 Randomisation

3.6.1 Clusters and stratification

Scoping work by MHCLG conducted in the lead up to the design of the CBEL trial indicated that some learners attended English classes with friends or family members, and being able to attend the course with a known family member or friend was important to their participation. On this basis, it was decided that individual level randomisation was not appropriate as it posed a risk of splitting groups of known trial participants (the consequences of which were anticipated to include greater recruitment challenges, participant retention issues and possible control group contamination).

The trial therefore adopted a cluster randomisation approach. All participants were asked at registration if they knew anyone else attending the course in that or another centre. Where this occurred, the names and unique learner identification reference numbers of all associates were noted on participants' registration forms. Participants known to each other were then randomised into the treatment and control conditions together.[13]

Around thirty per cent of trial participants belonged to a cluster larger than a single individual. The mean cluster size was 1.21 participants; for the treatment group the mean cluster size was 1.15 and for the control it was 1.28. Overall there were fewer clusters of learners within the treatment group with 76.7 per cent of learners registering alone compared to 63.7 per cent of the control group (χ^2_1=10.59, p=.001). It is possible that treatment group learners may be disadvantaged by not being able to benefit from out of course peer support that control group participants may have; however the implications of this are unclear. Table 1 shows the frequency of participant clusters by size.

---

[13] The registration form and associated Excel spreadsheet accommodated up to six associates to be added to an individual learner.

Table 1: Participant clusters by frequency and number of trial participants

Size of cluster	Frequency of cluster	Number of participants	Per cent of all trial participants
6 participants	2	12	2.3
5 participants	2	10	1.9
4 participants	3	12	2.3
3 participants	7	21	4.0
2 participants	52	104	19.7
1 participant	368	368	69.8
Total learners	527	527	100.0

Randomisation was further restricted by the need to ensure roughly equal numbers of participants were assigned to each of the trial conditions at each centre. This was important in order to minimise variation in class to tutor ratio (twelve learners to one tutor was considered the optimal balance). Therefore, randomisation was stratified by centre.

3.6.2 Implementation

All eligible trial participants who registered on the CBEL course were randomised into control or treatment group.

Randomisation only occurred once a centre had registered a sufficient number of participants to run two courses (one immediately for those assigned to the treatment group and one in September for those in the control group). Once achieved, the registration data for all participants in a centre were quality assured on site by MHCLG staff before being transferred via a secure online file transfer site to Learning and Work's London office for randomisation.

The randomisation was conducted in the statistical computing environment 'R' using anonymised data comprising participants' unique learner identification reference, cluster identifier (indicated by a single unique learner identification reference designated as a 'cluster lead' and attributed to all participants belonging to the cluster), and centre location. Randomisation was staggered as and when recruitment numbers were sufficient to do so at each location. In total, five discrete randomisation runs were conducted between late April and early May.

Once randomised, lists indicating the allocation of participants to the treatment or control condition were returned to the relevant provider centre. Treatment group participants were identified by an 'April/May' start date while control group participants were identified by a 'September' start date. Providers then contacted participants (by phone and letter) to inform them of their start date and invite them to an assessment session.

3.7 Blinding

Due to the nature of the intervention whereby the provider could identify a participant's condition assignment based on when they were due to attend the course, it was not possible to blind participants nor the provider to the allocation in this particular trial.

Moreover, as many of the baseline and follow-up assessment events separated treatment and control participants (to avoid the risk of them sharing learning/resources) this meant that neither assessors from ESB, nor researchers from BMG Research, were blinded to trial participants' assignment.

3.8 Sample size

Power calculations carried out prior to the implementation of the trial (detailed in the RCT protocol in Annex A) suggested that a final achieved sample of 400 participants equally split over both arms would be required (with power set at 0.8, and significance at 0.5). An estimated 30 per cent attrition rate was applied with an assumption that the mean 'known participant' cluster size would be 1.05. As such, the target sample size to be achieved at registration was set at 600 individual trial participants.[14]

Therefore 25 centres were identified by the course provider to participate in the trial, with each expected to recruit a minimum of 24 trial participants (to be split between the intervention and control groups).

Due to some difficulties in recruitment two centres failed to recruit sufficient numbers of trial participants to feasibly run the intervention, while some other centres over-recruited (for more detail see the accompanying process evaluation).[15] Overall 528 participants were recruited and randomised following the registration stage to 23 classes, across 22 centres. One trial participant was excluded from the trial after attending treatment condition despite being assigned to the control group reducing the effective sample size to 527. Given the effect size, this was a large enough sample to achieve statistical power.

Following randomisation, 249 trial participants were assigned to the treatment condition and 278 to the control. Figure 3 shows participant numbers at the point of registration, baseline assessment and at follow-up.

Overall 354 trial participants provided follow-up measures accounting for 62.1 per cent of all participants registered. 326 participants provided measures at both baseline and follow-up, accounting for 61.9 per cent of all trial participants. The attrition rate between baseline and follow-up (e.g. the proportion of participants who provided baseline but not follow-up measures) was 24.2 per cent. There was a lower rate of attrition among those in the treatment group (19.3 per cent compared to 28.5 per cent among the control group). This was not surprising given the increased contact between the treatment group and the provider.

---

[14] Revised power calculations conducted following receipt of baseline measures (retaining the same power and significance levels) suggested that in order to detect a mean difference in proficiency equivalent to a one point change in score between the two arms of the trial, would require a sample of 105 randomisation units, or 128 individual trial participants.

[15] https://www.gov.uk/government/publications/community-based-english-language-programme-a-randomised-controlled-trial--2

Figure 3: Participant flows through the CBEL RCT*
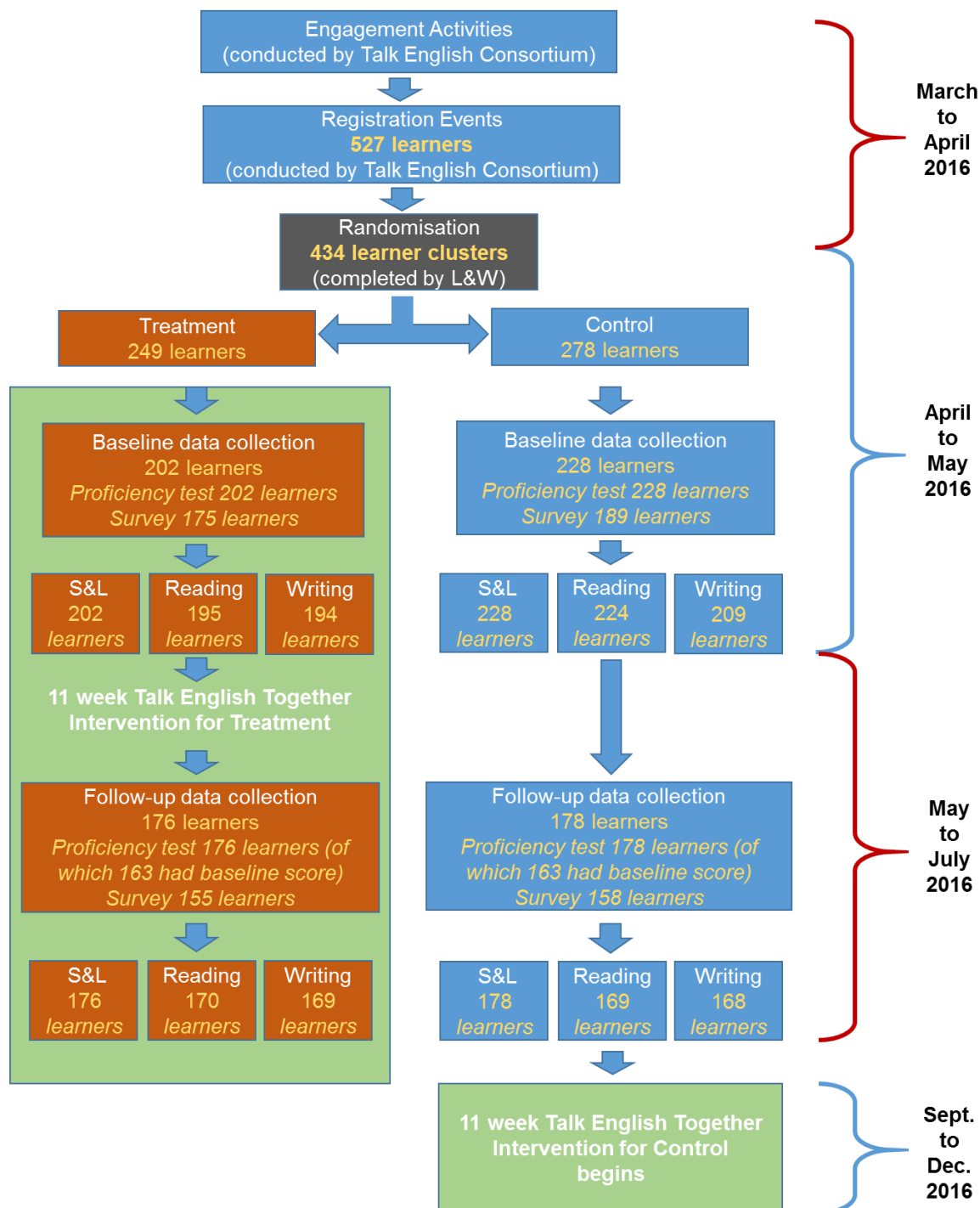
* This figure excludes 15 learners' reading and writing assessment data at a single centre (further details provided in section 3.9 below). Baseline assessments for a further four learners were excluded from analysis, as their unique identifiers were incorrectly recorded at this stage. Finally, one learner was also excluded from the trial entirely due to registering at two separate centres and being assigned to different trial conditions at each with the participant attending the CBEL intervention. Individual data on dropout was not collected throughout the trial.

3.9 Data

Analysis presented within this report is based on the following data sources:

English Proficiency Assessments: Data on English proficiency was collected by assessors from ESB with the support of CBEL tutors. The data was drawn from assessments administered at, or shortly after the start of the treatment intervention (either during the first session or within a week of the course start), and again on completion of the course for all participants. Assessments covered three domains of English proficiency: speaking and listening, reading and writing, each of which was assessed against a ten-point scale. Within each of these domains, three levels of assessments were developed: Pre-entry, Entry level 1 and Entry level 2. Table 2 shows the capability levels with the equivalent ESB English proficiency assessment score.

Participants scoring in the range of zero to three were considered to be Pre-entry level; those scoring four to six, were considered to be Entry 1, and those scoring seven to nine, were considered to be Entry 2. Follow-up assessments mirrored the baseline scoring framework, with a maximum ceiling score of 9 (Entry level 2) on any assessment.

Table 2: Expected capabilities corresponding to ESOL proficiency levels

Level	Equivalent ESB English proficiency assessment score	Speaking and listening (example capabilities)
Very limited (Pre-entry)	0-3	Low level of understanding; may be able to answer questions for basic personal information and follow basic instructions.
Basic (Entry 1)	4-6	Able to ask and answer basic questions for personal information in full sentences and give and follow basic instructions and directions. Answers may include errors, but the overall message should be understood.
Advanced beginner (Entry 2)	7 – 9	Can make simple statements, descriptions, requests and ask permission and for clarification. Can listen and respond to spoken language including straightforward requests, information and narratives, and follow straightforward explanations and instructions. Can engage in discussion with one or more people to communicate feelings and opinions on a familiar topic. Uses simple present, present continuous and simple past tenses, and can indicate future meaning, although may make grammatical errors, particularly in negative and question forms.
Level		**Writing (example capabilities)**
Very limited (Pre-entry)	0-3	May not be able to write anything or may be able to write some personal words and numbers.
Basic (Entry 1)	4-6	May be able to write a few short, simple sentences or phrases (would be able to compose very simple text to communicate ideas or basic information such as filling in a limited number of personal details on a form).

Advanced beginner (Entry 2)	7 – 9	Learners will be able to use written words and phrases to record or present information. They will be able to construct simple and compound sentences (containing two or more pieces of information) and link these by using connectives e.g. as, and, but and to use adjectives and punctuation correctly (capital letter at start of sentence, full stop at end). Uses simple present, present continuous and simple past tenses, and can indicate future meaning, although may make grammatical errors particularly in negative and question forms.
Level		**Reading (example capabilities)**
Very limited (Pre-entry)	0-3	May not be able to read anything or may be able understand basic signs and symbols in texts such as public signs and notices
Basic (Entry 1)	4-6	Can read and understand short texts with repeated language patterns on familiar topics. Can read and obtain information from common signs and symbols in texts such as public signs and notices, lists, forms and simple narratives. They will be able to recognise the letters of the alphabet in both upper and lower case.
Advanced beginner (Entry 2)	6-9	Can read and understand short, straightforward texts on familiar topics. Can read and obtain information from short documents, familiar sources and signs and symbols in texts such as public signs and notices, lists, forms, notes, records, e-mails, simple narratives, letters and diagrams. Can read and understand words on forms related to personal information, e.g. first name, surname, address, postcode, age, date of birth.

Pre-entry level assessments across all three domains were administered by English Speaking Board (ESB), as were the speaking and listening Entry level 1 and Entry level 2 assessments. CBEL tutors administered the delivery of the Entry level 1 and 2 reading and writing assessments, which were completed in group settings.

There was one centre which departed from this approach, whereby all assessments levels across all domains were carried out by the ESB assessors. This applied to the treatment and the control group equally, and was therefore not felt to introduce any bias or error into the data collected.

All of the assessments were marked by ESB assessors. English proficiency assessments took longer than had been originally anticipated. Despite this full measures were captured for the clear majority of participants. There were however, isolated incidents reported within centres of individuals having to leave before completing the full reading and writing assessments. Data gathered as part of the process evaluation indicated that these were infrequent occurrences affecting very few learners. This was often due to existing commitments (e.g. child care responsibilities, hospital appointments, etc.) and was reported to affect treatment and control group participants equally. In these instances, the affected individual was invited back to complete the outstanding assessments in a further baseline or follow-up event (where possible).

The duration of the assessment process did however result in one centre failing to administer any Entry 1 or Entry 2 reading and writing assessments to control group participants at baseline. While all of these individuals were invited to attend further

assessment sessions, very few did so. Reading and writing proficiency data from this centre were therefore excluded from the final analysis to remove any potential systematic bias.

Following this issue coming to light, extensive checks were carried out to ensure assessments at other centres were carried out as intended. Materials and communications were reviewed and further detailed enquiries were made with the provider in each of the five local authority areas to explore whether this or similar issues had occurred elsewhere. These inquiries suggested that this issue was localised to the particular centre and did not warrant the exclusion of other participants or centres from the analysis. In addition, there were no such issues identified at the follow-up point.

Further details about the implementation of the intervention, including the issues encountered with administration of the outcome measures, are included in the accompanying process evaluation.[16]

Social integration survey: A survey of trial participants was conducted before (or shortly after starting) and at the end of the course. The survey measured a range of social integration indicators and measures of personal independence. The survey used binary, categorical and ordinal (Likert scale) measures.

Only participants who spoke one of the specified languages (Arabic, Bengali, Punjabi, Somali or Urdu) were required to complete the survey. Overall, 364 learners (69.1 per cent of all trial participants) were able to complete the survey at the baseline.

The survey was administered on-site by BMG researchers with the support of additional interpreters fluent in one of the five core languages. Trial participants who were literate in one of the core languages were given the opportunity to self-complete the survey, though interpreters remained on-hand to provide support should it be required. The BMG researchers were careful to monitor participants to identify anyone who appeared to be struggling by providing *ad hoc* translation of the questions. Participants, who were not functionally literate were supported by BMG researchers who orally translated the survey on a one-to-one basis. Care was taken to ensure that there were sufficient numbers of BMG researchers and interpreters available at both the baseline and follow-up sessions to support all of the trial participants eligible to complete the survey to do so.

Administrative data: A range of socio-demographic data was captured as part of the registration process (through paper registration forms) including: the preferred centre, contact details, age/date of birth, gender, religion, family structure, education, motivation to attend and details about individuals' native and second languages, as well as other known learners. This data was inputted electronically into a standardised Excel spreadsheet by area coordinators. Data inputting was checked on-site by MHCLG staff, and once checked was forwarded via a secure online file transfer site to L&W's London office for randomisation. Paper registration forms were forwarded to BMG Research for entry onto a separate database which was subject to a full quality checking procedure.

This demographic data was supplemented by detailed lesson records, which allowed tutors to capture any unusual events (for example, low class attendance due to a religious

---

[16] https://www.gov.uk/government/publications/community-based-english-language-programme-a-randomised-controlled-trial--2

festival) reasons and mitigation for departing from the course material, and class attendance records, providing the number of classes treatment group participants attended.

3.10 Statistical methods and analysis

A number of different analyses were carried out to explore the key outcome measures in the trial. Descriptive and bivariate analysis was conducted in SPSS and Stata, while multilevel modelling was conducted in R using the 'lme4' package.

3.10.1 Analysis of English language proficiency

Analysis first sought to test the following primary hypotheses, measuring the level of change in English proficiency between the baseline and follow-up measure:

- H_{0a} = there is no difference in the level of improvement in the English language proficiency between the treatment and control groups; and

- H_{1a} = there is a difference in the level of improvement in the English language proficiency between the treatment and control groups

Analysis also explored the absolute difference in the post-intervention English proficiency scores. This analysis sought to test the following hypotheses, using an outcome measure based on the follow-up measure alone:

- H_{0b} = there is no difference in the observed levels of English language ability between the treatment and control groups after the treatment group receive the intervention; and

- H_{1b} = there is a difference in the observed levels of English language ability between the treatment and control groups after the treatment group receive the intervention

A set of linear multilevel models were fitted to data, to explore the relationship between condition assignment (control vs. treatment) and English proficiency outcomes.

Multilevel models are commonly used to correctly model the hierarchical structure of the dataset (Carvajal et al 2010, Goldstein 2011). In this case a three-level model was fitted to reflect the structure of data, whereby individual trial participants (Level 1), were nested within learner clusters (Level 2) which in turn were nested within centres (Level 3).

In this study, learners registered and signed up at the centre of their choice, and some also signed up with friends or family as a cluster. As a result, it is possible that different centres within the trial had different learner profiles. Similarly, it is also possible that different clusters of learners had different ways of responding to the CBEL programme. The multilevel modelling is therefore important as it controls for any effects that being in a particular cluster or centre may have on the trial outcomes. This is important, as there are

a number of consequences associated with not doing so (Rasbash et al 2009, Wears 2002) including underestimation of standard errors associated with model coefficients.

The multilevel model also controlled for any fixed effects (i.e. variables introduced into the model as an independent explanatory or predictor of outcome). The fixed effects used in the model include: condition of assignment, baseline score on the same domain, baseline after course start, native language, age, gender, length of residence in UK, prior education attainment, existence of children aged under 5 and existence of children aged 5-18.

Condition assignment was introduced into the model as the key predictor of interest, as it would be the key indicator of a causal relationship between the intervention and any changes observed.

Baseline score was introduced into the model to control for the relative English proficiency of trial participants at the start of the intervention. The variable *Baseline after course start* was introduced to reflect the fact that some among the treatment group had attended between one and three classes prior to baseline measures being obtained. While this may be an insufficient period of time to make substantive progress with regards to English proficiency, the inclusion of this explanatory variable was designed to control for any distortions that may have been created.

The remaining socio-demographic variables were introduced into the model in order to explore whether they were predictors of progress in English proficiency (outside of condition assignment). Model estimates provided for an individual predictor variable control for all other predictors introduced in the model. For example, the estimate provided for condition assignment takes into account the influence that all other variables (e.g. baseline score, age, gender, etc.) introduced into the model may exert.

The outcome variables for the models presented in this study included:

- change in English speaking and listening scores between baseline and follow-up;

- follow-up English speaking and listening scores;

- change in English reading scores between baseline and follow-up;

- change in English writing scores between baseline and follow-up;

- follow-up English reading scores;

- follow-up English writing scores;

The models take the form:

$$y_{ijk} = \beta_0 + X_{ijk}\beta_1 + v_{0k} + u_{0jk} + e_{ijk}$$

Where y_{ijk} is either the change in score or follow-up score depending on the run of the model for the ith individual in the jth learner cluster, in the kth centre; β_0 is constant term equal to the average across all groups; X_{ijk} is the covariate matrix for the fixed effects (noted above) β_1; v_{0k} represents the level 3 (centre) residuals; u_{0jk} the level 2 (learner

cluster) residuals; and e_{ijk} is the individual level error term clustered around centre and learner group.

As noted previously, all participants from one centre were excluded from the analysis of reading and writing outcomes. This was due to control group participants within that centre not being fully assessed at Entry level 1 and Entry level 2 on both the reading and writing assessments. Individuals within this centre were however fully assessed on the speaking and listening domain and therefore not excluded from these analyses.

To reflect the exclusion of this particular centre from some analyses, findings relating to the speaking and listening scores are presented separately from that relating to the reading and writing assessments.

The models used a modified "intention to treat" (ITT) approach, meaning they considered the likely impact of the intervention on all those who entered the programme and for whom baseline measures are available, regardless of whether or not they completed the programme. Excluding individuals that fail to complete the course from analysis of outcomes would skew the estimates upwards, particularly if participants that make, on average, less progress are more likely not to complete the course.

ITT analysis allows for estimation of the effect of being assigned the intervention, even where that individual does not provide a follow-up measure. To account for missing data at the follow-up stage, follow-up measures are computed on the basis of a last observation carried forward (e.g. baseline scores uplifted by any positive change in the scores of the control group). This approach makes the following assumptions about treatment group participants who do not complete the follow-up measures due to dropping out or non-attendance:

1. Their scores will be less than those who stay on the course and do complete follow-up measures;

2. Their scores will not be lower than any positive change observed in the control group.

Overall, using an ITT approach provides a more conservative measure of impact on English proficiency and improves the reliability of impacts identified. This analytical approach is also appropriate for assessing the effect of a social intervention in which one could expect a relatively high degree of dropout or attrition. This approach provides a more 'real world' understanding of impact, as it includes all those starting the intervention (rather than only those who provided follow-up measures).

Model outputs for both the difference in speaking and listening score and follow-up speaking and listening score are presented below. All other model outputs can be found in Annex E.

Estimates and standard errors in the models can be interpreted in much the same way as for standard single-level regression models. Each explanatory variable has a reference category; for binary measures, this was taken as zero. However, for non-binary categorical variables the reference category was selected by default in the R package as the *first* category within a variable; for example, in the case of language, each group is compared to Arabic speakers (the reference category). Reference categories in the tables can be

identified by having an estimate of zero, and no standard error. Positive estimates indicate an increase in score compared to the reference category, while negative estimates indicate a decrease.

When interpreting model outputs, it is important to note that the estimates provided are **not standardised** and therefore should not be used to rank different predictor variables (e.g. to identify the most important predictor) as the units of measure across the variables are not the same. Further explanation of this is provided in the results chapter, when presenting the model outputs.

Model outputs presented in this report focus on the speaking and listening assessment, though key statistics from models fitted to the reading and writing outcome variable are also presented, along with commentary on notable differences in the models. Full model outputs for reading and writing are presented in Annex E.

3.10.2   Analysis of the social integration survey

The CBEL intervention was primarily designed to achieve an improvement in English language proficiency. As a consequence, the improved English language proficiency was expected to improve capacity to mix and integrate in wider society. Analysis of the survey data explores the extent to which participating in a Community-Based English Language course impacted on integration over the 11-week duration of the course.

Analysis of survey data compared differences in responses between treatment and control groups, applying appropriate tests of statistical significance to identify noteworthy differences. Analysis specifically sought to test the following hypotheses:

- H_{0c} = there is no observable difference in the levels of social integration or activity between the treatment and control groups; and

- H_{1c} = there is an observable difference in the levels of social integration or activity between the treatment and control groups.

Throughout the presentation of findings significant results are indicated within tables and charts by the presence of an '*'. Levels of significance are indicated in the following way: * = significant at the 5% level; ** = significant at the 1% level; and *** = significant at the 0.1% level.

4 Results

4.1 Condition balance

Ensuring a balance between treatment and control groups is important in any RCT, as it is through the comparison of these similar groups that differences can be attributed to the intervention. While the two groups are unlikely to be identical (due to random variation) checking the balance between the groups (by looking for systematic differences) indicates the nature and extent of these variations.

This section presents analysis on the balance of the trial arms, first in terms of the primary baseline measures around English proficiency and secondly the socio-demographic characteristics of the trial participants.

In order to test the balance in terms of baseline score, mean scores were compared between the treatment and control groups using an independent *t-test.* With regards to socio-demographics individual characteristics were compared using *Pearson's chi squared test.*

4.1.1 Baseline measures

Checking the balance in English proficiency test scores between the treatment and control groups at the baseline stage found no statistically significant variation on speaking and listening or writing measures. Baseline scores for speaking and listening were 4.42 for the treatment group compared with 4.27 for the control group (t=0.768, p=.443). Baseline writing scores were 3.63 for the treatment group and 3.45 for the control group (t=0.893, p=.372).[17]

There was, however, a significant difference in reading scores between groups at baseline with mean scores varying from 4.95 for the treatment group, to 4.40 in the control group (t=2.315, p=.021). It should also be noted that unlike speaking and listening and writing scores, which have a normal distribution, the distribution of reading scores appears to be bimodal.[18] Figures 4 and 5 show speaking and listening and writing scores following a normal distribution, while reading scores for both the treatment and control groups appear to have a bimodal distribution, with peaks at three points, and six points, with a third peak appearing at the nine-point mark.

Reasons for the imbalance in reading scores are not clear, however one possible explanation, which could also explain the unusual distribution of scores found in both the control and treatment groups, relates to measurement bias. The modal peaks at three, six and (less markedly) nine points on the scoring scale correspond to ceiling scores for Pre-entry level, Entry level one and Entry level two English proficiency scores. This scoring pattern could indicate a tendency among assessors to mark at these specific cut points. However, such patterns were not evident in writing scores, which were assessed in the

---

[17] Due to control group participants in one centre not being fully assessed in reading and writing, data for this centre was excluded from analysis related to these domains.
[18] Differences between reading scores in the treatment and control groups were nevertheless significant when considered using a statistical test appropriate for non-normally distributed data (treatment median = 5.5, control median = 4, U(430) =20886.0, p<0.05).

same way. No additional evidence was found to confirm the existence of measurement bias.

Other possible explanations include 'ascertainment bias', which refers to bias introduced as a result of not being able to blind the ESB assessors and tutors responsible for administering the reading tests, to participants' trial assignment. However, ascertainment bias does not appear to have influenced baseline speaking and listening and writing scores, the latter of which also involved CBEL course tutors in its implementation so its influence is likely to be low.

It should be noted that tests were not conducted in exam conditions. It is therefore possible that trial participants may have collaborated during the course of the assessments. While it would be harder to collaborate on speaking and listening and writing assessments, reading may have been more prone to this type of bias.

However, data collected as part of the process evaluation did not substantiate any of these possibilities. It is therefore not possible to determine whether the differences observed between treatment and control reading observations could be attributed to other, unobserved causes.

In any case, the difference in baseline reading does not appear to be related to the initial randomisation of trial participants and while the difference should be noted, it is reassuring that scores in the other English proficiency domains (i.e. writing and speaking and listening) are not significantly different. Importantly, the statistical modelling applied to assess the impact of the intervention on English proficiency, controls for differences in baseline score. This reduces any distortion created by the difference in baseline reading between the trial and control groups, therefore providing a degree of confidence in the findings.

Figure 4 Distribution plots of TREATMENT group baseline speaking and listening, reading and writing scores*
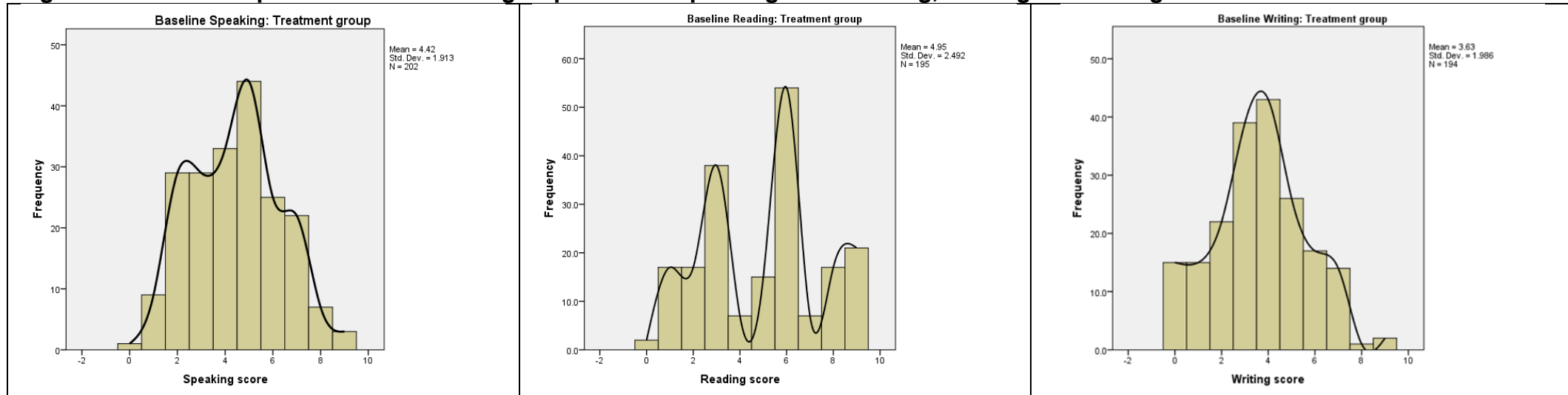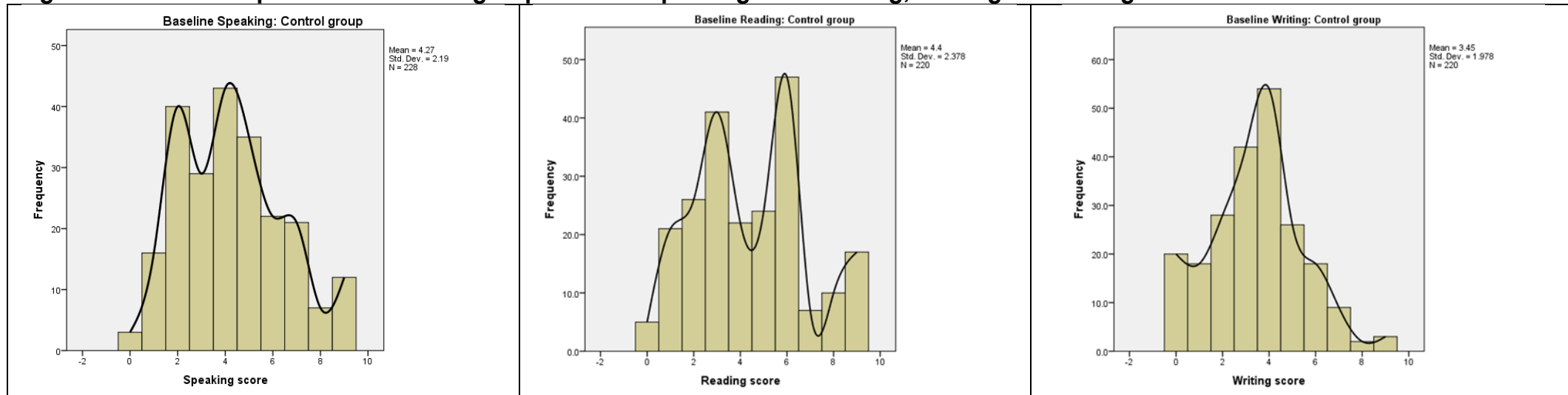

Figure 5 Distribution plots of CONTROL group baseline speaking and listening, reading and writing scores*

* Baseline reading and writing scores exclude learners from one centre where full measures were not obtained from the control group.

4.1.2 Socio-demographic characteristics

In order to check that treatment and control groups are comparable (at least on observable variables), demographic balance checks were conducted at three stages:

1. at the point of randomisation (for all registered learners);
2. attendance at baseline measures session (for those who attended the baseline session);
3. attendance at follow-up session event (for those who attended the follow up event).

The results of these checks can be seen in Table 3. Overall, both groups are well-balanced. While there are some differences observed (outlined below) they do not appear to be systematic. A full presentation of trial participant socio-demographics can be found in the results section. Differences include:

o At the point of randomisation, the only imbalance detected was in the make up of languages spoken (there was a significantly higher proportion of Bengali speaking participants in the control group).

o Though not significantly different at the point of registration, at the baseline measures stage, significant differences were noted in the proportion of participants with secondary school education (with higher rates being reported in the treatment group) and length of time living in the UK (significantly more treatment group participants had lived in the UK less than 3 years).

o At the follow-up measures stage, again the only signficant difference between the characteristics of learners in the treatment and control groups was in the make up of languages spoken (the control group included a a significantly higher proprtion of Bengali speakers).

Overall, socio-demographics were similar across the treatment and control groups, and the differences detcted do not appear to be systematic or overly influential on outcomes (see results, below).

Table 3: Demographic balance at randomisation, baseline and follow-up points

Group		Randomisation			Baseline			Follow-up		
		n	Treatment (%)	Control (%)	*n*	Treatment (%)	Control (%)	*n*	Treatment (%)	Control (%)
Gender	Female	454	85.89	87.64	378	85.44	86.32	306	88.57	85.8
	Male	69	14.11	12.36	62	14.56	13.68	45	11.43	14.2
Age	18-24	33	6.85	5.82	29	8.25	5.11	23	8.00	5.11
	25-34	175	35.08	32	146	33.98	32.34	121	36.00	32.95
	35-44	175	32.66	34.18	143	31.07	33.62	113	31.43	32.95
	45-55	105	18.15	21.82	91	19.42	21.7	70	17.14	22.73
	56+	35	7.26	6.18	32	7.28	7.23	24	7.43	6.25
Language	Urdu	143	28.92	25.54	122	28.99	26.16	104	31.82	26.97
	Arabic	33	6.02	6.47	25	6.28	5.06	17	5.68	3.93
	Bengali	**60**	**8.03***	**14.34***	52	8.7	14.35	**42**	**6.82****	**16.85****
	Punjabi	50	8.43	10.43	43	7.73	11.39	35	8.52	11.24
	Somali	3	0.8	0.36	3	0.97	0.42	3	1.14	0.56
	Other	237	47.79	42.45	198	47.34	42.19	152	46.02	39.89
Area	Oldham	88	16.87	16.55	71	16.91	15.19	55	12.5	18.54
	Rochdale	45	8.03	8.99	44	9.66	10.13	38	10.8	10.67
	Kirklees	106	20.48	19.78	76	15.94	18.14	57	17.05	15.17
	Manchester	175	32.93	33.45	153	34.78	34.18	130	37.5	35.96
	Bradford	113	21.69	21.22	100	22.71	22.36	74	22.16	19.66
Education	NFE	148	26.51	29.5	123	25.12	29.96	90	26.7	24.16
	Primary	127	23.69	24.46	106	23.19	24.47	89	24.43	25.84
	Secondary	148	31.73	24.82	**130**	**34.78***	**24.47***	107	32.95	27.53
	College	83	13.65	17.63	69	13.04	17.72	54	11.93	18.54
	University	19	4.42	2.88	14	3.86	2.53	12	3.98	2.81
Years in the UK	0-2 Yrs	173	36.95	29.14	**155**	**40.1***	**30.38***	128	39.77	32.58
	3-4 Yrs	55	11.24	9.71	48	11.11	10.81	34	9.09	10.11
	5-9 Yrs	115	19.82	24.1	92	16.91	24.05	77	19.32	24.16
	10-14 Yrs	64	12.05	12.23	50	12.08	10.55	35	10.8	8.99
	15-24 Yrs	54	8.84	11.51	46	9.18	11.39	37	8.52	12.36
	25+ Yrs	60	10.04	12.59	50	10.14	12.24	40	11.93	10.67
Children under 5	Yes	304	68.9	69.57	259	68.42	71	215	70.2	74.66
Children over 5	Yes	193	52.46	49.24	164	51.3	50.9	131	53.91	51.67

4.2 Outcomes and estimation

This section presents the results of the comparison of English proficiency and social integration outcomes between treatment and control groups.

Speaking and listening outcomes are presented first, followed by summary results for proficiency in reading and writing (full results for reading and writing are included in Annex E).

Proficiency outcomes are considered in two stages.

- Firstly, comparison of the mean (unadjusted) changes in score and the mean (unadjusted) follow-up scores observed in treatment and control groups. These analyses provide an indication of the impact of the intervention on English proficiency.

- Secondly, the application of statistical models to identify the key predictors of changes in score and follow-up scores. These analyses offer a further indication of the impact of the intervention on proficiency, taking account of other factors that could have an impact on English proficiency.

Finally, this section presents findings of the analysis of the social integration survey.

4.2.1 Speaking and listening proficiency

4.2.1.1 Change in English speaking and listening proficiency

Attending the intervention course had a significant impact on the level of change detected between the treatment and control groups in the English speaking and listening proficiency assessment.

Comparing the mean difference in the change in score between speaking and listening measures showed statistically significant improvements in the treatment group. As can be seen in Figure 6, the level of improvement in speaking and listening score within the treatment group was around double that achieved by the control group (a 1.39 increase in score compared to a 0.71 increase ($t(324)=3.653$, $p<.001$) respectively).

Figure 6: Mean change in speaking and listening scores between baseline and follow-up by trial condition
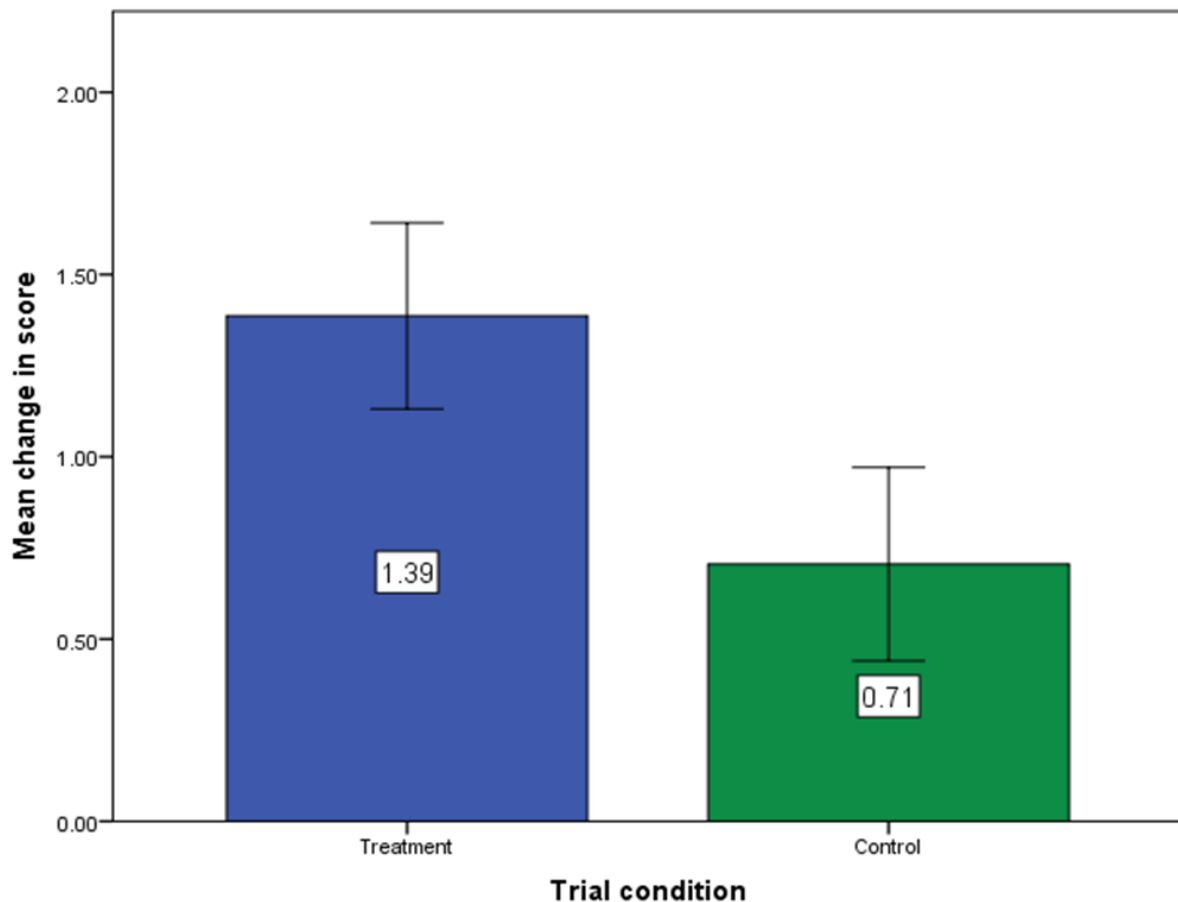
Interestingly, when looking at the repeat measures over time, the changes observed in both the treatment and control conditions were positive and highly significant. All trial participants experienced a significant improvement in English proficiency, irrespective of their assignment into the treatment or control group[19]. However, as presented above, the level of improvement within the treatment group was significantly greater than that found within the control group.

An accompanying process evaluation exploring implementation of the RCT found little evidence of contamination within the control group (i.e. control group learners having been exposed to the intervention). As such, contamination is not considered a plausible explanation for the improvement in English proficiency observed among the control group[20]. However, the process evaluation did find that the control group altered their learning behaviour in response to registering on the CBEL course (to prepare for the course in September). These changes found among the control group are therefore likely to be influenced by a trial effect (commonly referred to as the Hawthorne effect). This refers to the process by which participants' behaviour

---

[19] The test statistics for comparing mean follow-up speaking and listening scores against baseline scores was highly significant for the treatment group (t(162)=10.733, p<.001), Likewise, it was highly significant for the control group (t(162)=5.245, p<.001).
[20] https://www.gov.uk/government/publications/community-based-english-language-programme-a-randomised-controlled-trial--2

changes as a consequence of being enrolled onto the trial itself. Whether an improvement in English proficiency would have occurred had participants not registered for the Talk English Together course or entered into the RCT is not known.

Comparing the follow up English speaking and listening proficiency scores for treatment and control groups also found a strong and sizeable difference between the groups. As can be seen in Figure 7, the mean follow-up score for speaking and listening is significantly higher for the treatment group compared with the control group. The treatment group scoring, on average, almost a full point higher than the control group (respectively means were 5.91 compared to 4.92; $t(351)=4.928$, $p<.001$).

Figure 7: Mean follow-up speaking and listening scores by trial condition
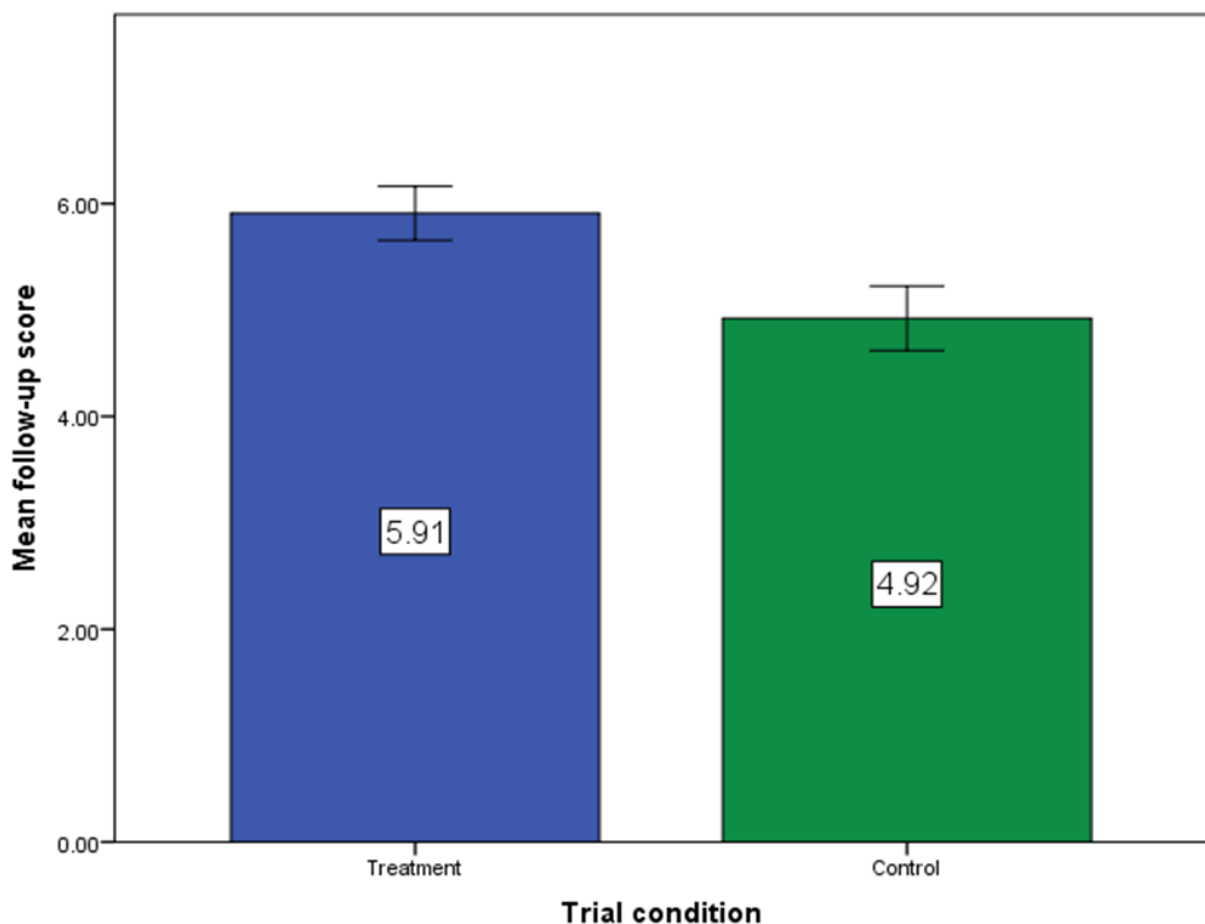
4.2.2 Predictors of change in speaking and listening score

To explore the factors which contributed to performance in the assessments, a linear mixed model was fitted to the data. Within this we controlled for variables that may have affected the level of change in score between measures. Table 4 presents the model output (further outputs can be found in Annex E). All significant predictors of the change in English proficiency score are highlighted in bold with an indicator of the level of significance.

Estimates and standard errors can be interpreted in much the same way as standard single-level regression models. For each predictor, the model estimate indicates the size of the effect on the outcome variable compared to the reference category (identified by an estimate equal to zero and a missing standard error).

Trial condition (i.e. receiving the CBEL intervention), further education, having a child over the age of 5 and baseline English speaking and listening score were all significant predictors of the change in English speaking and listening proficiency scores.

Being in the treatment group was a significant predictor of change in English speaking and listening scores (compared to the control group). Controlling for all other variables in the model, the change in the treatment group participants speaking and listening scores was 0.92 points (t=3.70, p<.001) higher than the control group.

Controlling for other variables, model estimates suggest that the control group (as the reference category) would improve in their speaking and listening by 1.41 points (in line with the model intercept). The treatment group (again controlling for all other variables) are predicted to improve their score by an additional 0.92 points, meaning an overall improvement of 2.33 points. This figure is obtained by adding the estimate for the treatment group to the constant value represented in the table as the 'intercept'. The level of improvement expected in the treatment group is 65.2 per cent higher when considered relative to change seen in the control group who, controlling for all other variables in the model, would be expected to improve 1.41 in line with the constant (intercept) value alone.

Table 4: Multilevel model of change in English speaking and listening score between assessment waves

Predictor	Group	Estimate	Std. Err	t-value
(Intercept)		1.41	0.56	2.53
Condition	Control	0.00	-	-
	Treatment*	**0.92**	**0.25**	**3.70**
Baseline test score	**English speaking***	**-0.30**	**0.05**	**-6.68**
Language	Arabic	0.00	-	-
	Bengali	-0.31	0.43	-0.73
	Other	-0.37	0.40	-0.94
	Punjabi	-0.10	0.46	-0.21
	Somali	-1.64	1.38	-1.83
	Urdu	-0.07	0.40	-0.17
Gender	Female	0.00	-	-
	Male	-0.49	0.26	-1.85
Age group	18-24	0.00	-	-
	25-34	0.35	0.36	0.98
	35-44	0.22	0.38	0.57
	45-55	-0.11	0.43	0.26
	55+	-0.09	0.48	-0.18
Education	No formal education	0.00	-	-
	Primary	0.20	0.22	0.92
	Secondary	0.12	0.23	0.55
	College	0.05	0.29	0.19
	University**	**1.68**	**0.64**	**2.62**
Time in UK	Years	0.04	0.06	0.60
Child under 5	No child u5	0.00	-	-
	Child u5	0.18	0.19	0.96
Child 5 or over	No child 5/+	0.00	-	-
	Child 5/+*	**0.45**	**0.18**	**2.49**
Baseline after course start	No	0.00	-	-
	Yes	-0.25	0.27	-0.96

* p<.05; ** p=<.01; *** p<.001

With regards to education, when compared with the reference group (no formal education) having a university degree increased the change in score by 1.68 (t=2.62, p<.01) when controlling for all other variables in the model. Likewise, having a child over the age of 5 increased the change in score by 0.45 (t=2.49, p<.05) when compared with those without children aged over 5.

It should be noted that though the model estimate for the treatment group at 0.92 appears relatively smaller than the estimate of 1.62 for those with a university education, it **does not** mean that it is a less influential predictor than university education. Estimates should only be considered relative to the reference category within the predictor variable it is nested in, so 'university' should only be considered relative to other groups within 'education' predictor (e.g. no formal education,

primary, etc.). They cannot be directly compared to estimates for other predictor variables.

With regards to baseline score, the level of change decreased as baseline scores increased (indicated by the negative value of the model estimate); that is to say, for every one-point increase in the baseline score, the change in score between the measures decreased by 0.3 (t=-6.68, p<.001).

Figure 8 below shows the predicted change in speaking and listening score by baseline score and condition assignment based on model estimates, controlling for all other variables included in the model.

As is clear from Figure 8, the treatment group achieved significantly greater changes (i.e. larger improvements) in speaking and listening proficiency compared to the control group. However, the graph also illustrates that there was a decline in the magnitude of change in proficiency across both groups as baseline scores increased (e.g. those higher up the scale at baseline achieve smaller improvements in proficiency over the course of the trial).

Figure 8: Predicted change in speaking and listening score by baseline and condition assignment based on model estimates
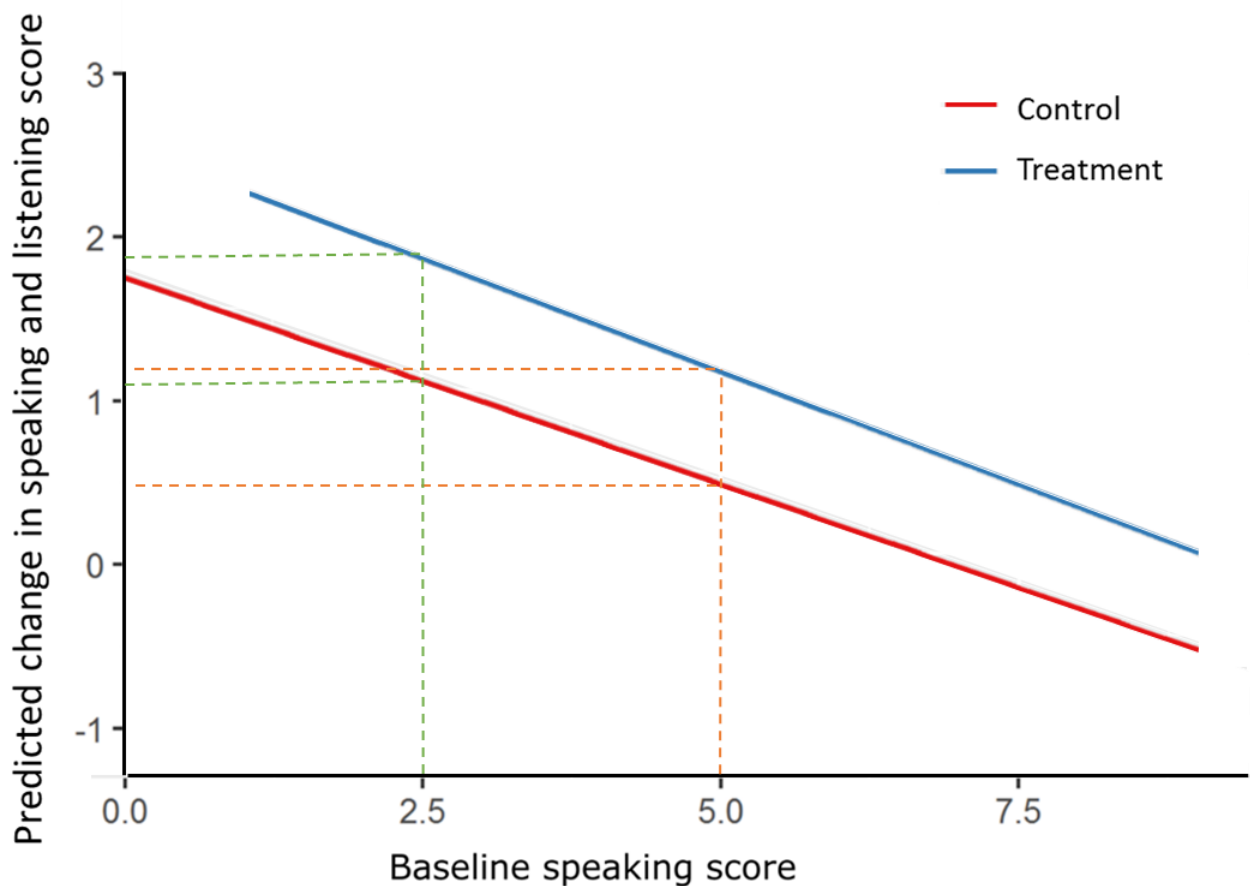
As can be seen from Figure 8, participants in the treatment group scoring 2.5 at baseline were predicted to achieve an almost 2 point improvement in their score at the follow-up point (indicated on the chart with a green dashed line). However, those

scoring 5 at baseline were predicted to achieve just over one point improvement (indicated on the chart with the orange dashed lines). Individuals scoring 7.5 at baseline were predicted a negative change in score. Similar patterns are evident among control group participants. The negative change in score among individuals scoring highly at baseline is also seen in analysis of the unadjusted data, as seen in Figure 9.

Figure 9: Mean change in English speaking and listening proficiency score
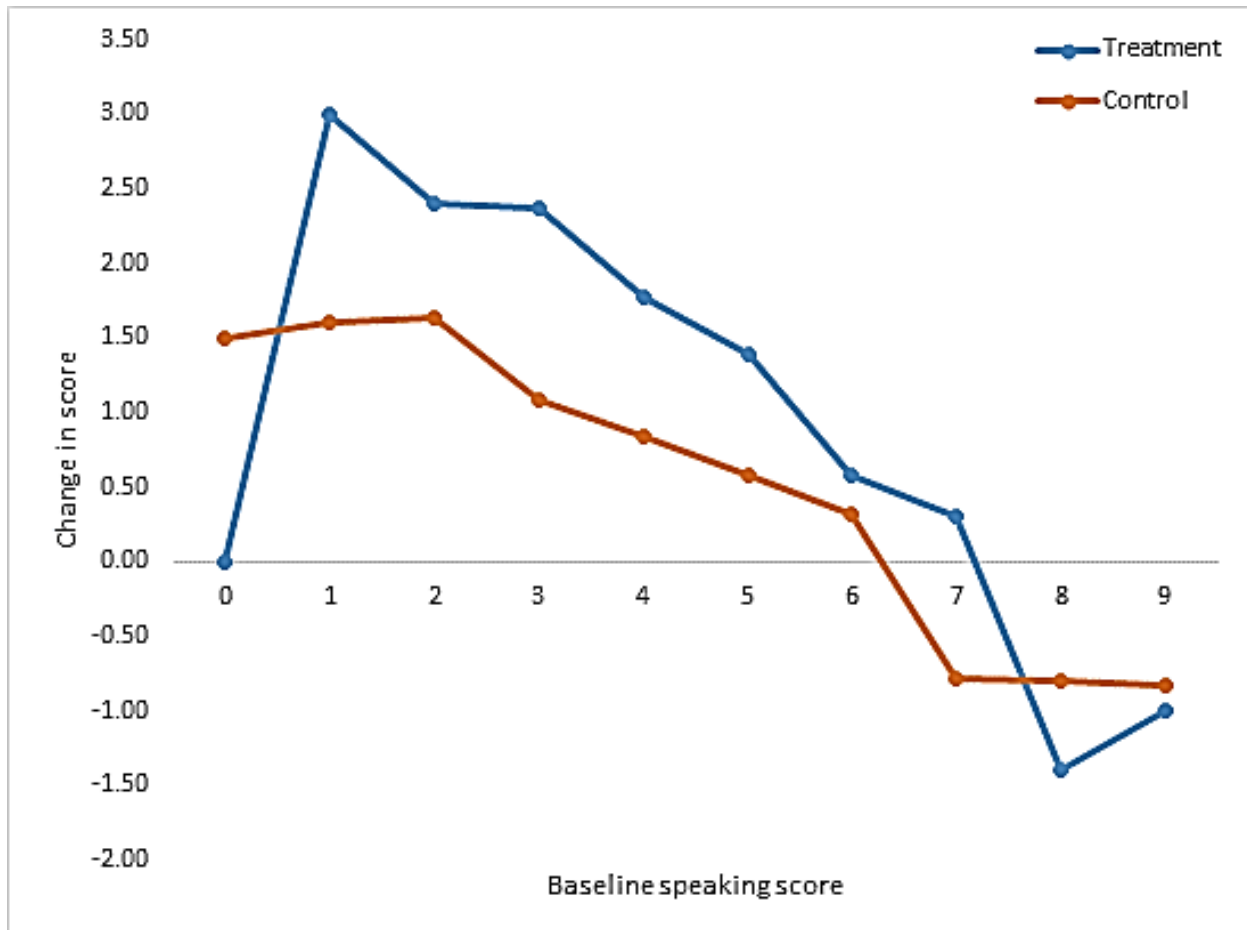
This is likely to be a function of the English proficiency assessments having a maximum score of 9. This effectively meant that those participants who scored relatively highly at the baseline had less scope within the assessment to progress, irrespective of any actual improvements made in their English proficiency. Overall, this is likely to have supressed the mean in the change in score between measures but will have affected both groups equally with the exception of the reading assessments (on which participants in the intervention group had higher proficiency at baseline). In this instance, the ceiling effect will have been greater for the intervention group.

4.2.2.1 Predictors of follow-up speaking and listening score

Table 5 presents the results for the linear mixed-effects model of follow-up speaking and listening scores. Controlling for all other variables introduced as predictors, the model shows trial condition, baseline score, university education, and having

children aged 5 and over were statistically significant predictors of the follow-up English speaking and listening score.

Model estimates suggest that treatment group participants scored 1.2 points higher on speaking and listening assessments at follow-up compared with the control group (the reference category) (t=3.65, p<.001).

With regards to baseline score, model estimates show that each additional point rise in the baseline score would lead to an increase of 0.49 in the follow-up speaking and listening score (t=8.04, p<.001).

Table 6: Multilevel Model of English speaking and listening follow-up score

Predictor	Group	Estimate	Std. Err	t-value
(Intercept)		1.97	0.55	3.59
Condition	Control	0.00	-	-
	Treatment*	**0.95**	**0.24**	**3.93**
Baseline test score	**English speaking***	**0.56**	**0.04**	**12.91**
Language	Arabic	0.00	-	-
	Bengali	-0.19	0.41	-0.51
	Other	-0.30	0.38	-0.76
	Punjabi	-0.02	0.45	-0.05
	Somali	-2.03	1.33	-1.53
	Urdu	0.02	0.38	0.07
Gender	Female	0.00	-	-
	Male	-0.45	0.25	-1.79
Age group	18-24	0.00	-	-
	25-34	0.25	0.35	0.72
	35-44	0.11	0.38	0.30
	45-55	-0.06	0.42	-0.14
	55+	-0.20	0.48	-0.41
Education	No formal education	0.00	-	-
	Primary	0.22	0.21	1.05
	Secondary	0.05	0.21	0.24
	College	0.17	0.28	0.62
	University**	**1.38**	**0.61**	**2.26**
Time in UK	Years	0.03	0.06	0.48
Child under 5	No child u5	0.00	-	-
	Child u5	0.13	0.18	0.68
Child 5 or over	No child 5+	0.00	-	-
	Child 5+*	**0.51**	**0.18**	**2.88**
Baseline after course start	No	0.00	-	-
	Yes	-0.15	0.26	-0.57

* p<.05; ** p=<.01; *** p<.001

Having university education is also positively correlated with follow-up English scores when compared with those with no formal education (the reference group) (t=2.82,

p<.01). The same is found with those who had a child aged 5 or older (t=2.81, p<.01). No other variables introduced as predictors reached statistical significance in explaining follow-up scores.

4.2.3 Reading and writing assessments

Analysis relating to both the reading and writing proficiency draws on data which excludes one centre due to inconsistencies in the way in which measures were obtained (see section 3.9).

4.2.3.1 Change in reading proficiency

As noted in section 1.2.4, balance checks identified statistically significant variation in reading scores between the treatment and control groups at baseline. As such the mean (unadjusted) follow-up reading scores should be treated with caution.

As with speaking and listening, there was a strong and significant difference in the change in the reading proficiency over time between treatment and control groups. The mean change being 1.65 and 0.69 points respectively (t(310)=3.619, p<.001) suggesting treatment group learners achieved more than double the level of improvement seen in the control group.

Interestingly, participants in both groups appeared to have improved their proficiency over time (test statistics comparing the baseline and follow-up means for the treatment and control group were respectively t(157)=8.741, p<.001, and t(153)=3.760, p<.001). Importantly however improvements among the treatment group were significantly larger.

Figure 10: Mean change in reading scores between baseline and follow-up by trial condition
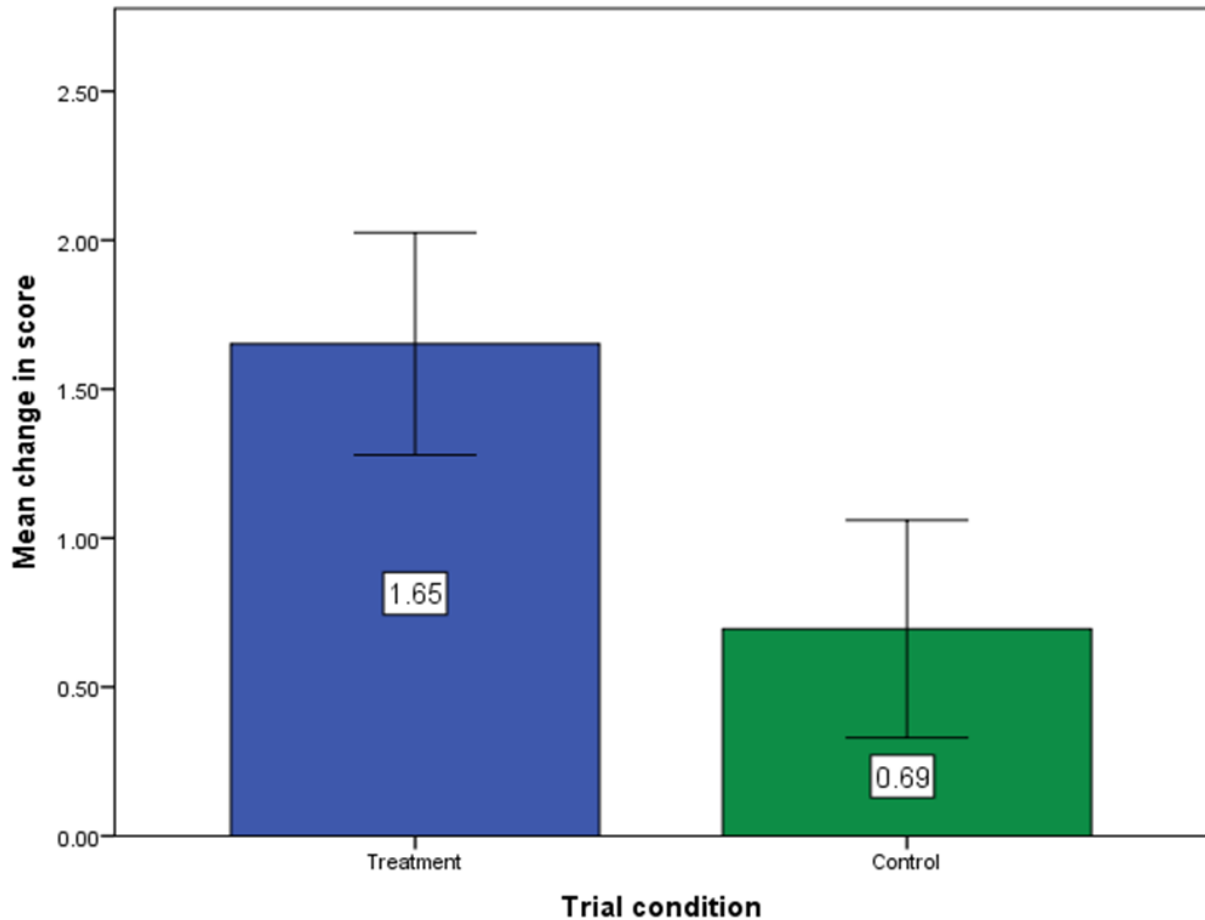
The mean difference in the score achieved at follow-up was also significantly higher among the treatment group (compared to the control group). As shown in Figure 11, the mean follow-up score for the treatment group was 6.65 compared to 5.09 for the control group (t(337)=5.237, p<.001). This result should however be interpreted cautiously given the significant differences in reading scores observed at baseline (see section 1.2.4). This finding alone should not be used to infer the intervention impacted on participants' reading proficiency, as differences in follow-up scores may simply reflect the greater existing reading proficiency of treatment group participants.

Figure 11: Mean follow-up reading scores between baseline and follow-up by trial condition
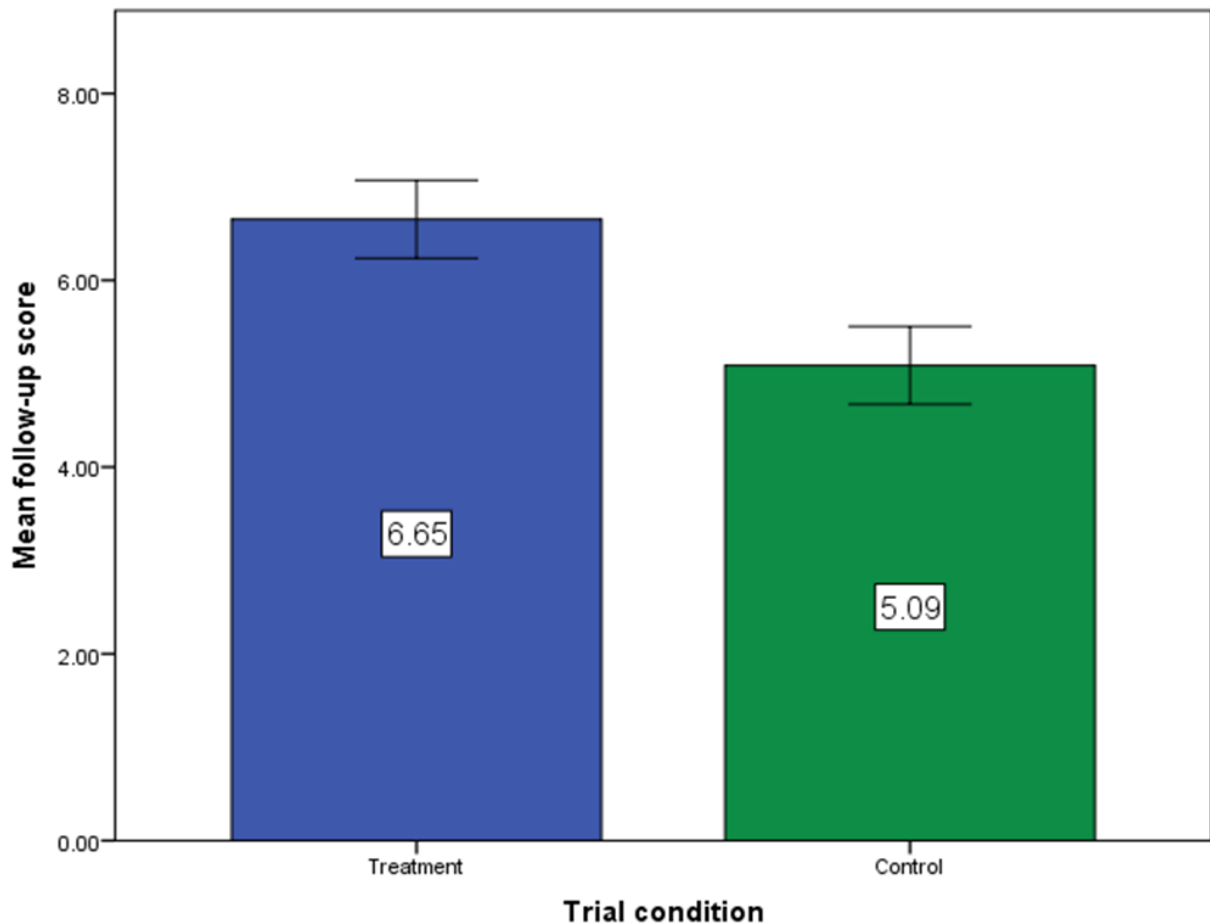
4.2.3.2 Predictors of change in score and follow-up reading score

In order to control for other factors that may affect reading proficiency linear mixed models were fitted to both the change in score and follow-up score data. These models also controlled for the imbalance of scores between the intervention and control groups.

Results showed that trial condition and baseline scores were both strong predictors of changes in reading scores over time. Language spoken, age and education and having children aged 5 and over were also predictors of changes in reading score.

Speakers of Bengali (compared to speakers of Arabic), those with both primary and college education (compared to no education) and those with children aged 5 and over all achieved greater improvements in reading score. Conversely, increasing age was associated with lower improvements in reading score (those aged 25-55 saw reducing changes in score compared to those aged 18-24).

Controlling for all other variables in the model, participants in the treatment group were estimated to score 1.29 points higher than the control group (t=3.42, p<.001).

Similarly, significant predictors included in the model of follow-up reading score were trial condition, baseline score, language spoken, age, education and having children

aged 5 or older. Participants in the treatment group were estimated to score 1.28 points higher than the control group at follow-up (t=3.42, p<.001).

Full model outputs can be found in Annex E.

4.2.3.3 Change in writing proficiency

Consistent with analysis of speaking and listening scores, there was a strong and significant difference in the change in writing proficiency over time both within and between the treatment and control groups.

As with the other English proficiency measures, there was a statistically significant improvement in writing proficiency across both treatment and control groups when comparing the baseline and follow-up means (respectively t(156)=8.917, p<.001); t(152)=4.103, p<.001).

Figure 12 shows the mean change in writing score was 1.46 points for the treatment group and 0.61 points for the control (t(308)=3.869, p<.001). This suggests that the treatment group achieved more than double the level of improvement in writing proficiency seen in the control group.

Figure 12: Mean change in writing scores between baseline and follow-up by trial condition
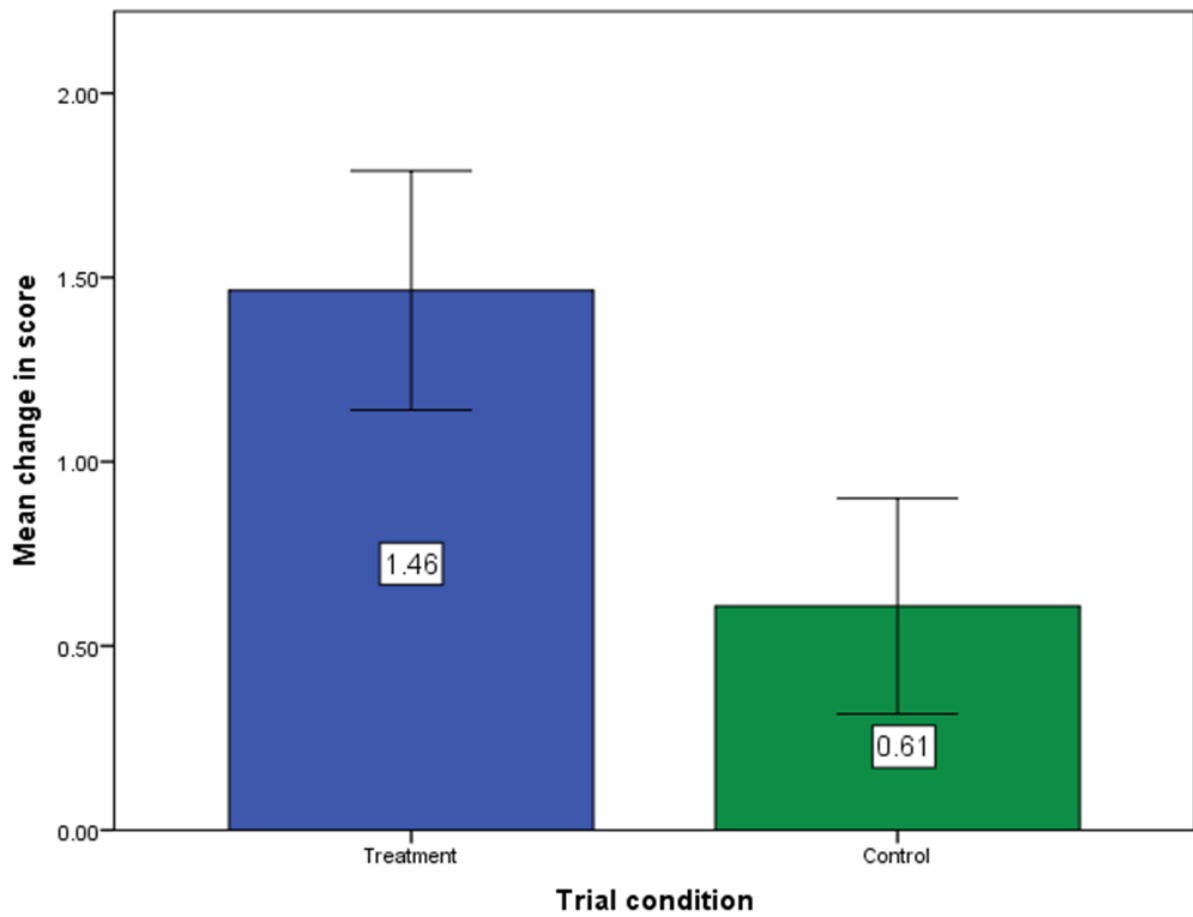
When comparing follow-up writing scores (see Figure 13) treatment group participants scored a full point higher than the control group (t(355)=3.935, p<.001).

Figure 13: Mean follow-up writing scores between baseline and follow-up by trial condition
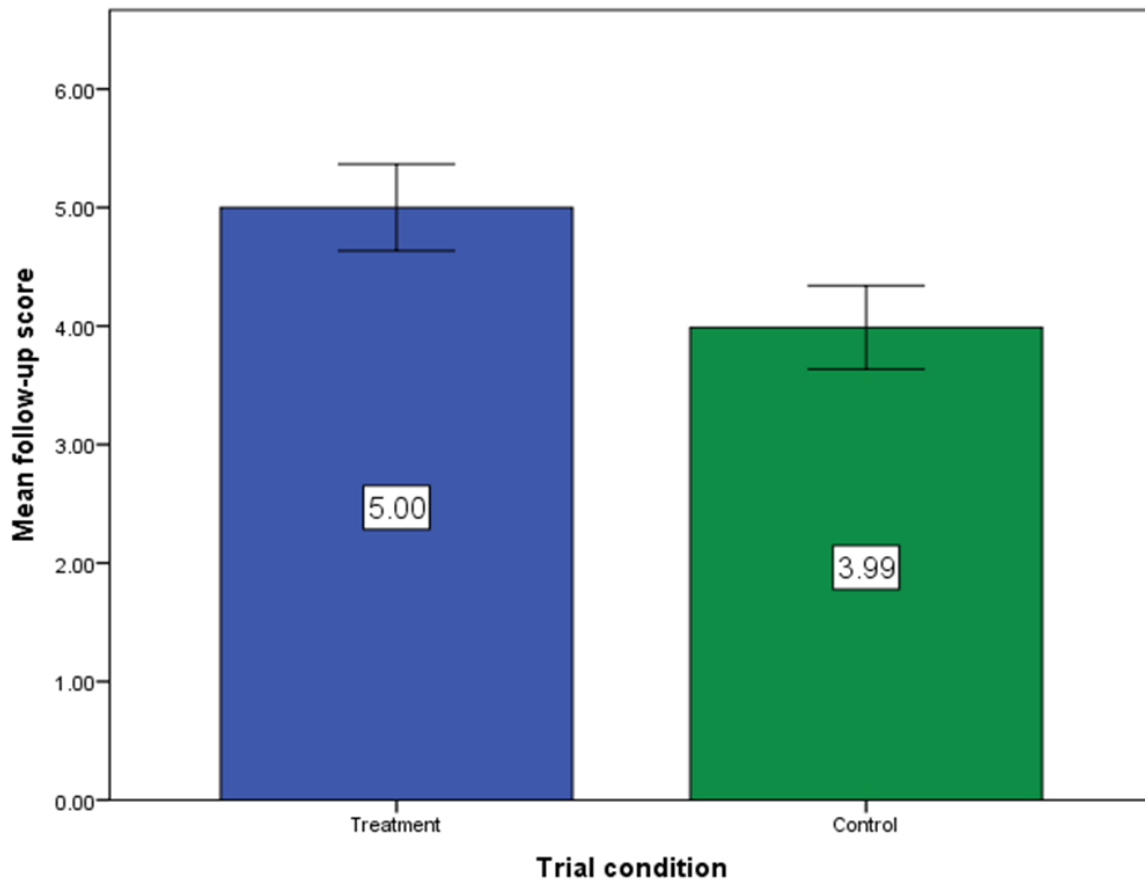
4.2.3.4 Predictors of change in score and follow-up writing score

The linear mixed model showed that trial condition and baseline scores were again strong predictors for changes in writing proficiency and the follow-up writing score. Language spoken, age and education were also predictors of both outcomes.

With regards to change in writing score (controlling for all other variables in the model) the treatment group was estimated to see an improvement of 0.90 points higher than the control (t=2.83, p<.01). Follow-up writing score was also estimated to be 0.90 points higher among the treatment group (again controlling for all other variables in the model) compared to the control group.

Full outputs for both models can be found in Annex E.

Analysis of attendance data, including exploration of the relationship between CBEL session attendance and English proficiency outcomes, can be found in the process evaluation.[21]

---

[21] https://www.gov.uk/government/publications/community-based-english-language-programme-a-randomised-controlled-trial--2

4.2.4 Social integration outcomes

This section presents analysis of the statistically significant differences observed between treatment and control group from the social integration survey. In line with the design of the survey itself, analysis is presented in several sections: social interaction; everyday activity; confidence in engaging with public services; independence; trust in others; community integration and belonging; and skills employment and training.

The impact of the CBEL intervention on social integration was mixed, with statistically significant differences detected on some questions but not others. Importantly, where significant differences were found at the follow-up stage, they were not significant at baseline. This suggests that the differences in results between groups emerged after the intervention, indicating a causal relationship between the specific survey measure and the CBEL intervention.

In order to test statistical significance, survey responses were aggregated into a binary form. For example, questions relying on a four-point confidence scale (typically, very confident, confident, not confident, not confident at all) were recoded into a binary form (confident, not confident).

Full survey output tables (including responses at baseline and follow-up) and details of the statistical tests undertaken can be found in Annex F.

Questions on social interaction and everyday activity explored the number of individuals and activities participants had engaged with over the last week. Analysis of this data explored whether there was an increase in reported interactions/activities between the two surveys. For example, whether there had been an increase in the number of conversations a respondent has had in English between the baseline and follow-up surveys.

Analysis of all other survey questions relating to confidence in engaging with public services; independence; trust in others; community integration and belonging; and, skills employment and training draw on data from the follow-up survey alone. Analysis of baseline survey data, which can be found in Annex F, showed no statistically significant difference between treatment and control group at baseline.

As such, changes observed at the follow-up stage can be plausibly attributed to the intervention.

4.2.4.1 Social interaction
Treatment group participants were significantly more likely to report an increase in the number of people they talked English to (outside of their English class) compared with the control group (χ^2_1=11.51, p=.001, 95% CI [-0.34, -0.09][22]). As shown in

---

[22] The CI represents the confidence interval in the difference between two proportions. In this instance the difference between the proportions reporting an increase is -0.21. The lower bound of the confidence interval is -0.34, while the upper bound is -0.09. The confidence level is set at 95%. Given this, the analysis suggests that one can be 95% confident between 9 and 34 per cent fewer control group participants would report increases in the number of people talked to in English when compared to the treatment group.

Figure 14, over 60 per cent of the treatment group reported more conversations in English post intervention (compared to 39 per cent in the treatment group).

Figure 14: Reported increase between baseline and follow-up – 'Apart from your English class, how many people did you speak to last week using English?'
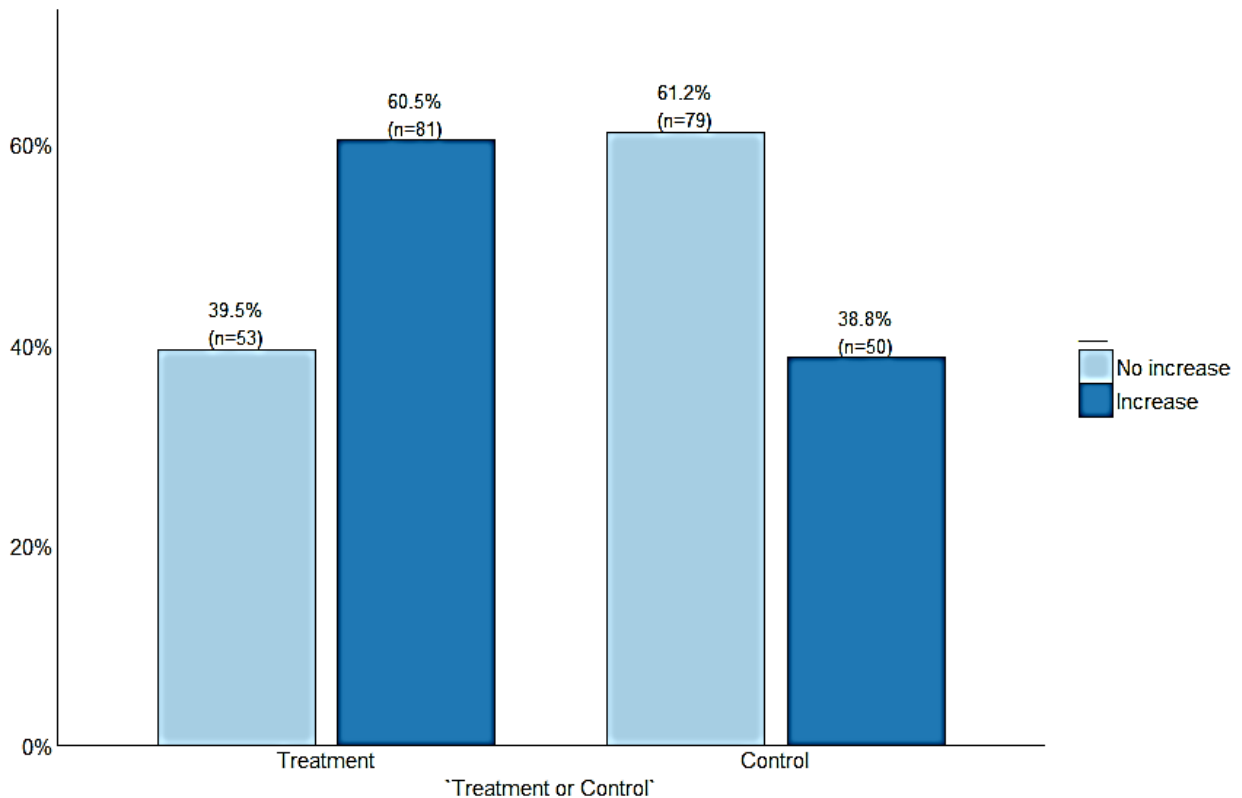
Treatment group participants were also significantly more likely to report an increase in the number of people talked to from a different background between the baseline and follow-up survey (x^2_1=11.84, p=.001, 95% CI [-0.37, -0.10])[23]. As Figure 15 indicates, nearly twice as many particpants in the treatment group (53.7 per cent) reprted having more contact with people from a different background post intervention compared to the control group (30.2 per cent).

---

[23] As noted previously, 51 respondents were excluded from the analysis of this question due to an error in the translation of this question into the Punjabi version of the translation. The incorrect survey question was applied to both the treatment and control group equally, and affected similar numbers across both groups, so did not unduly affect one group more the other.

Figure 15: Reported increase between baseline and follow-up – 'Apart from your English class, how many people did you speak to last week from a different country or religion to you?'
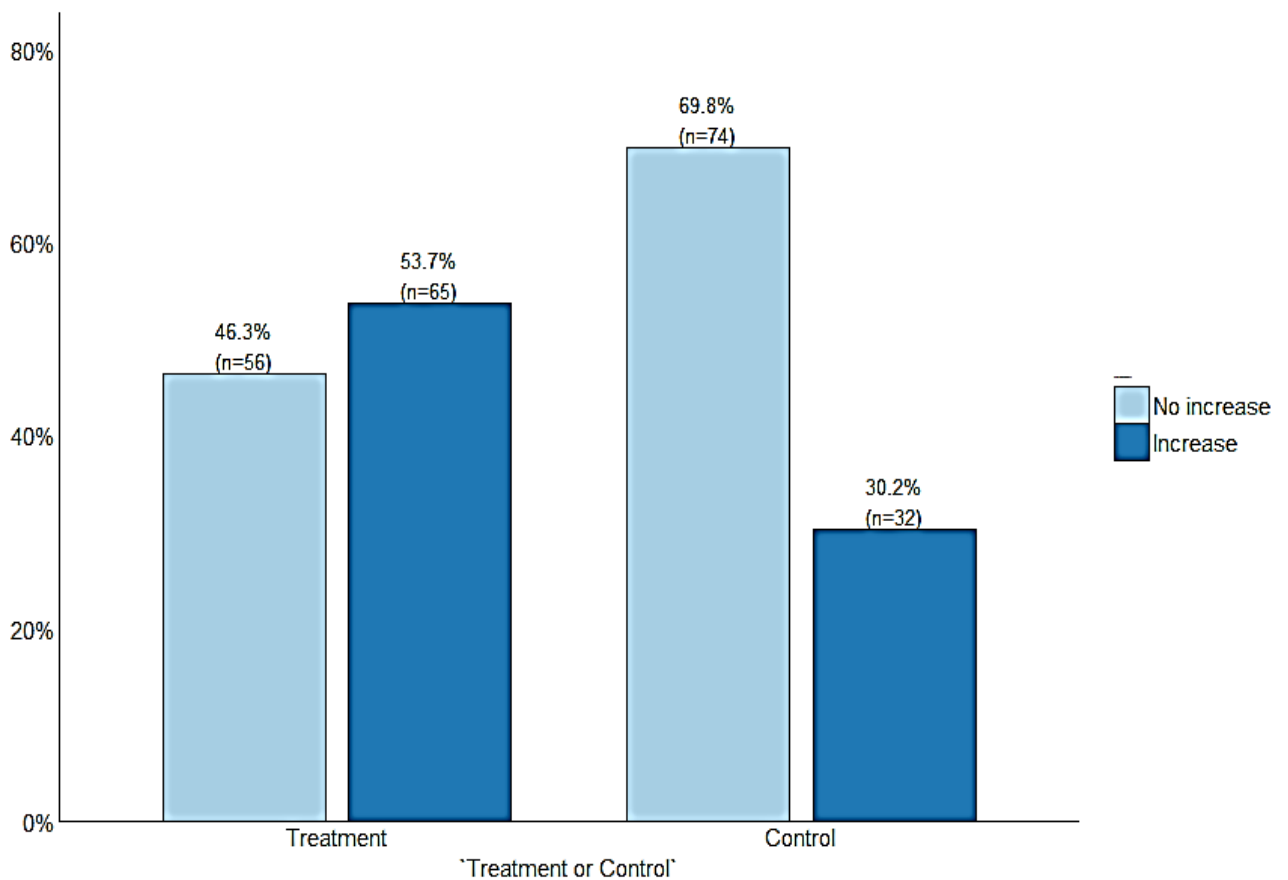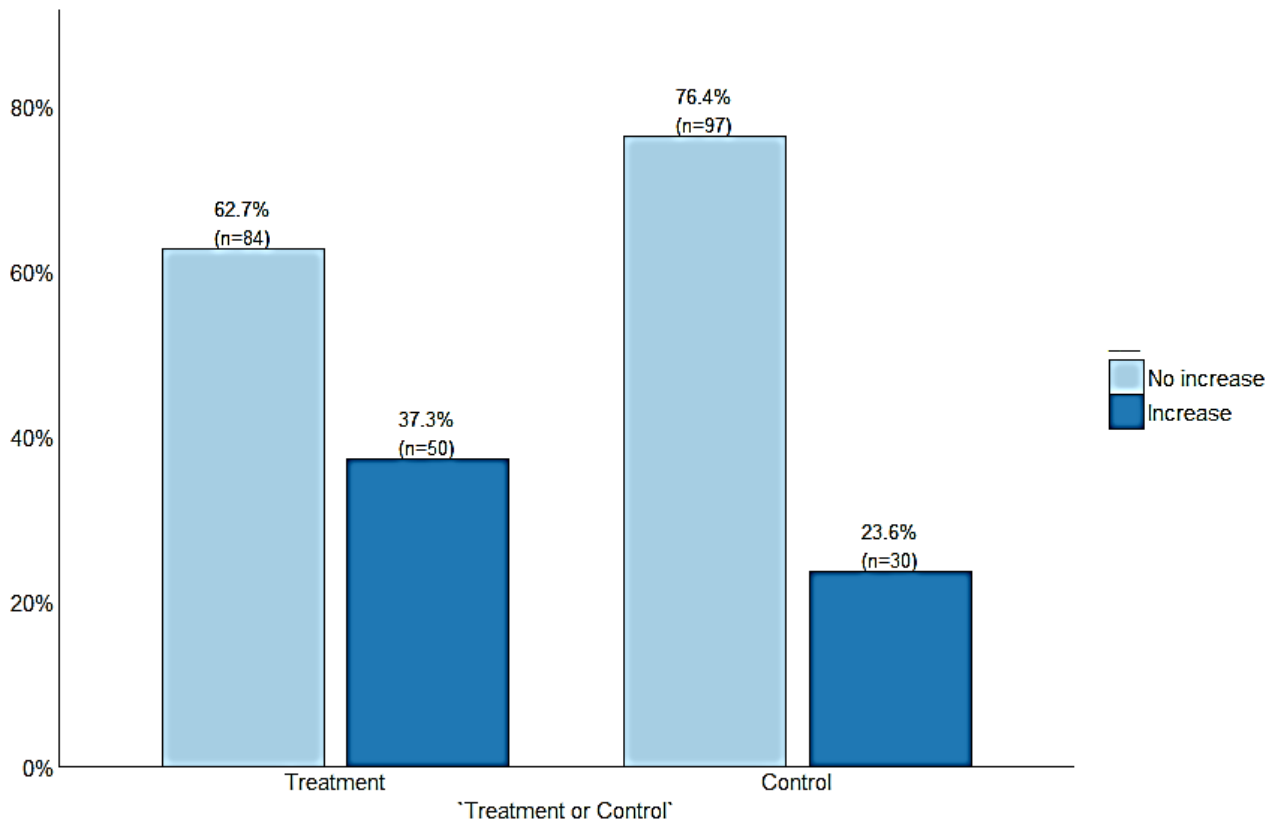

There was no evidence of the intervention having an impact on the number of friends respondents could rely on if they encountered a problem or required advice ($\chi^2_1=0.43$, p=.514, 95% CI [-0.16, 0.07]).

However, participants in the treatment group did report a significantly greater increase in their number of friends from a different background compared to the control group (Figure 16; $\chi^2_1=5.12$, p=.024, 95% CI [-0.25, -0.02]).

Over a third (37.3 per cent) of people in the treatment group reported having more friends from a different country or religion post intervention. This was compared to 23.6 per cent who reported an increase in friends from different backgrounds in the control group. This suggests that the intervention led to a significant increase in social mixing and bond forming with people from outside of trial participants' own cultural communities.

Figure 16: Reported increase between baseline and follow-up – 'How many of your friends are from a different country or religion to you?'


Overall there was strong evidence of the intervention increasing social interaction. All but one of the measures highlighted positive changes over the period of the trial, with increased interaction and bond forming. With regards to the lack of difference detected in the number of friends that an individual could rely on, one explanation could be that such relationships take longer to develop than was achievable between the taking of the measures.

4.2.4.2 Independent activity

With regards to everyday activities, participants in the treatment group were more likely to report an increase in going to the shops alone or without someone who speaks English (Figure 17; χ^2_1=3.10, p=.015, 95% CI [-0.28, -0.03]) than the control group. Half of those in the treatment group reported making more independent visits to the shops compared to 34.4 per cent of the control group.

Figure 17: Reported increase between baseline and follow-up – 'Apart from with your English class, how many times have you gone to the shops or market, either on your own, or without another person who speaks English in the last week?'
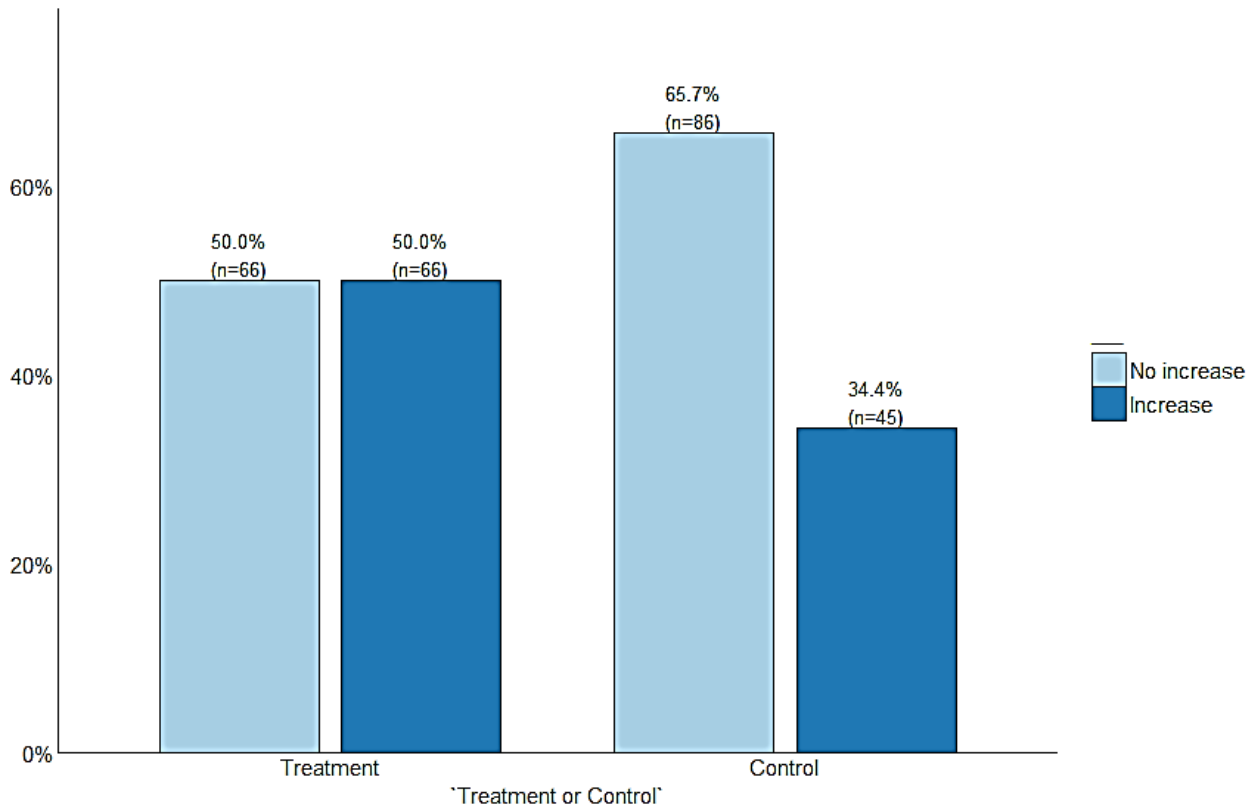
There were however no statistically significant differences detected in the use of public transport between the treatment and control groups (respectively 30.6 per cent versus 21.2 percent reported an increase; χ^2_1=2.58, p=.108, 95% CI [-0.21, 0.02]).

Similarly, no differences were detected in frequency of visiting parks or playgrounds between the treatment and control groups (respectively 42.5 per cent versus 32.1 per cent reported an increase; χ^2_1=2.68, p=.102, 95% CI [-0.21, 0.02]).

Overall, while the evidence around impact on everyday activity is less conclusive than in relation to social interaction, it does indicate that the CBEL intervention increased participants' independent use of shops and markets. This is promising and may be indicative of growing confidence in the independent use of local amenities. While it may have been expected that use of public transport would be related to an increase in visits to the shops, the fact that this is not seen may in part reflect the deeply localised communities respondents live in (which may not necessitate the use of transport).

4.2.4.3 Confidence in engaging with public services

As can be seen from Figure 18 control group participants felt significantly less confident talking to healthcare professionals than treatment group participants (χ^2_1=8.82, p<0.01, 95% CI [0.05, 0.25]). Rates of confidence among the two groups were similar at baseline, suggesting the greater confidence reported by the treatment group is a result of the CBEL intervention.

Figure 18: Reported at follow-up- 'How confident are you to book an appointment in English with a doctor, dentist, or nurse?'
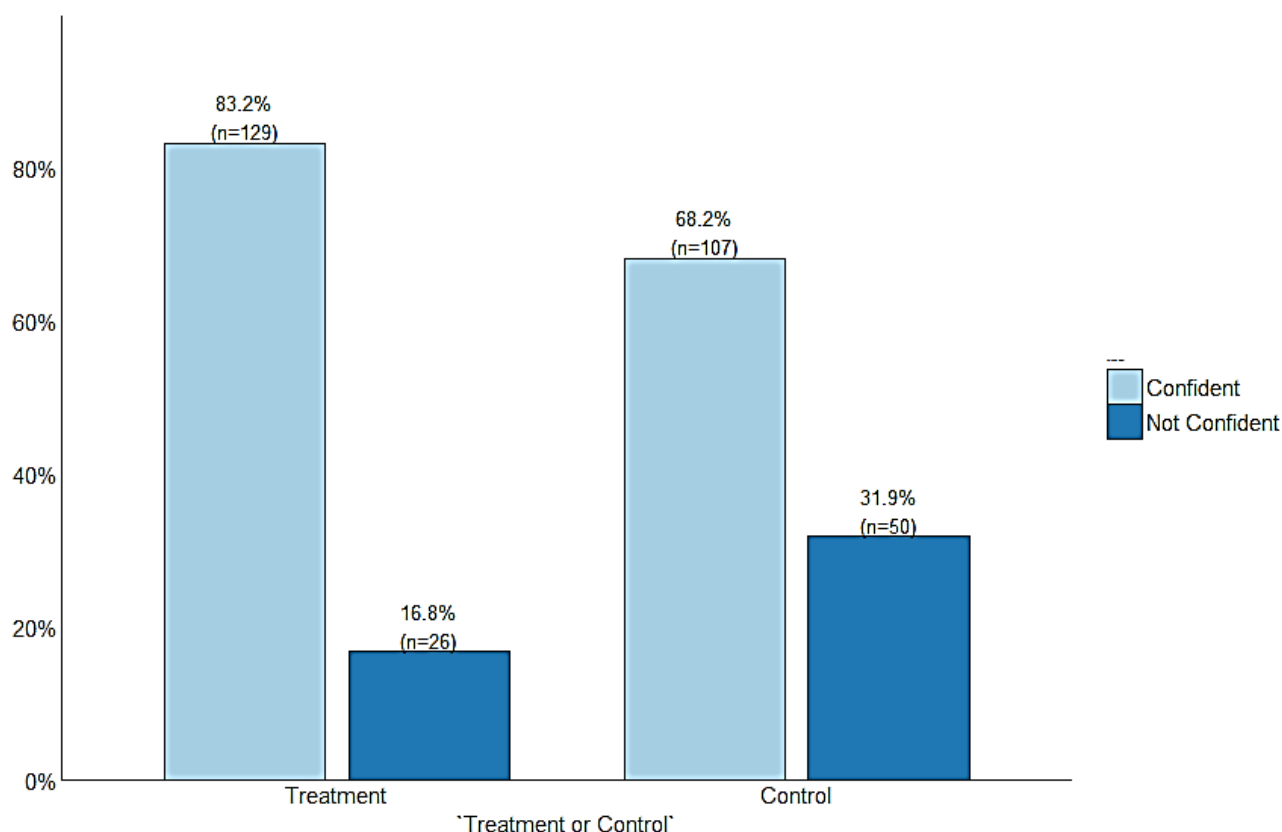
There were no statistically significant differences between the treatment and control groups in terms of confidence in talking to people at their child's school (respectively 76.0 per cent versus 63.6 per cent who reported being confident in talking to people at a childs school); χ^2_1=3.10, p=.078, 95% CI [-0.01, 0.26]) or to the police (68.2 per cent of the treatment group reported being confident compared to 59.0 per cent for the control; χ^2_1=2.45, p=.117, 95% CI [-0.02, 0.21]).

It is unclear why the intervention improved participants' confidence speaking to health professionals but not the other groups considered. One explanation may be the specific module on health within the intervention (delivered over 2 weeks) which included a visit from healthcare professionals. In relation to engaging with the police, most people could be argued to have a relatively low level of interaction with the

police and therefore this itself may make respondents less confident about this type of interaction.

4.2.4.4 Independence

There was no evidence that participating in the intervention had any impact on independence or personal agency. Overall very high proportions of participants across both treatment and control groups reported being able to decide how they lived their lives; respectively 91.6 per cent of the treatment group and 88.5 per cent of the control group reported being 'free to decide for myself how to live my life ($\chi^2_1=0.52$, p=.472, 95% CI [-0.04, 0.10]).

### 4.2.4.5 Trust and attitudes to social mixing

Levels of trust amongst one's own family were very high among both treatment and control respondents. However, participants in the treatment group were significantly more likely to report trusting their family members than the control group (Figure 19, Fisher's p=.043, 95% CI [0.001, 0.11]). It is unclear why this may be, but possible reasons may include improved communication within the home. Indeed, qualitative data collected as part of the process evaluation suggested that participation in the course had supported communication with others (usually children) within their household.

Figure 19: Reported at follow-up – 'How much do you trust people in your family?'
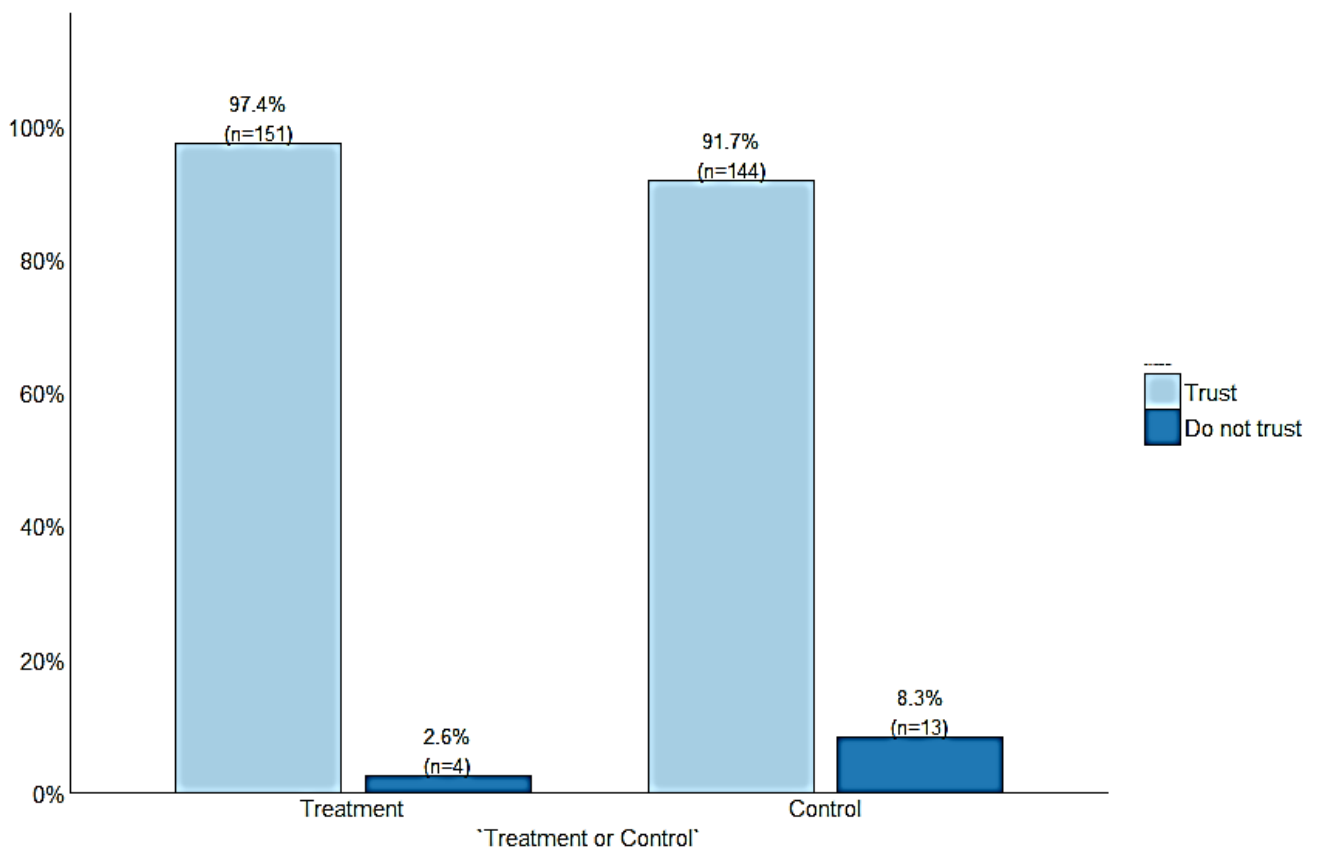
Similarly, as can be seen in Figure 20, levels of trust in the local community were significantly higher among participants in the treatment group (χ^2_1=13.16, p=.001, 95% CI [0.07, 0.24]). Differences in levels of trust in people from different countries or religions however fell short of statistical significance at the conventional 5 per cent level (92.9 per cent of treatment group and 77.6 per cent of control group participants; χ^2_1=13.16, p=.056, 95% CI [0.07, 0.24]).

Figure 20: Reported at follow-up - 'How much do you trust people in your local area?'
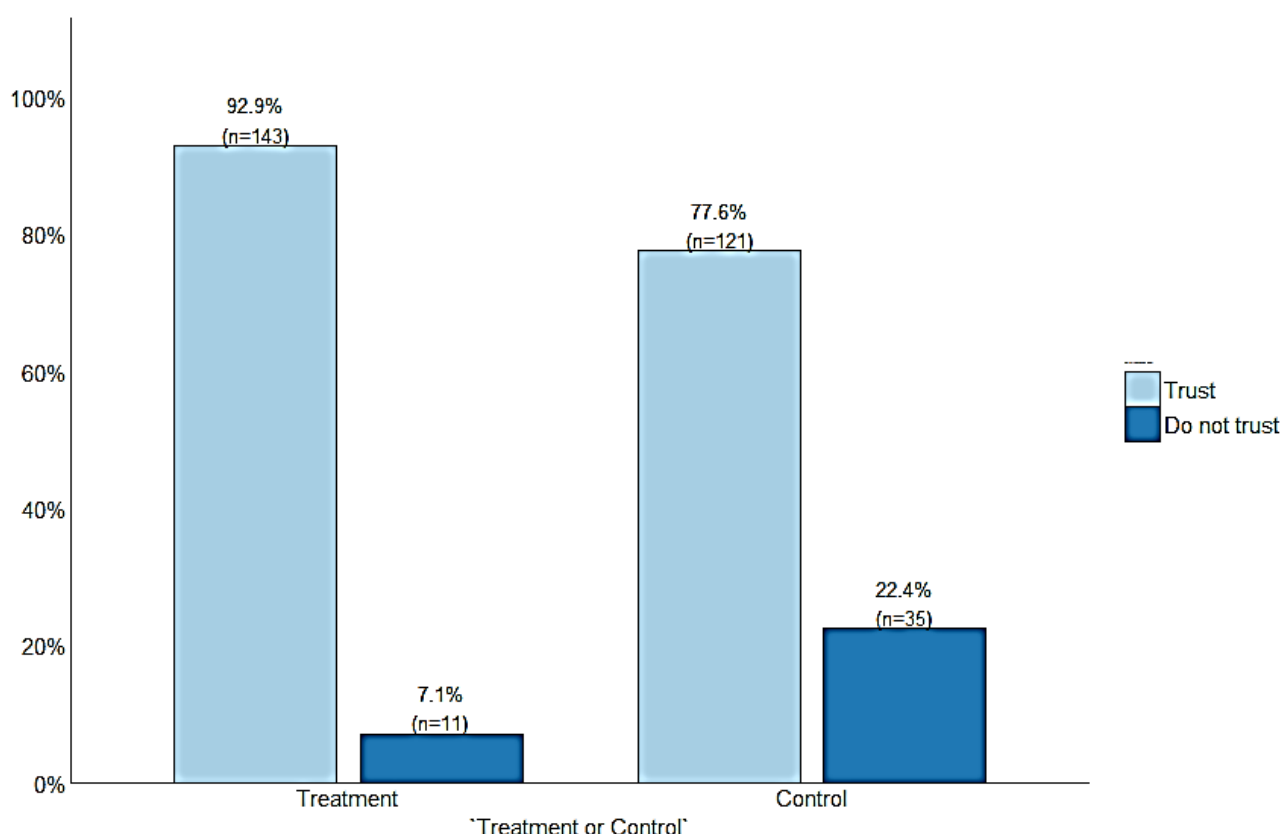
Participants were asked about their views on the level of mixing between people from different countries or religions, specifically whether they should 'mix less,' 'mix enough' or 'should mix more'. There were no statistically significant differences between treatment and control group responses (Mann-Whitney U=11862.5, Z=-.099, p=.921) with 25.3 per cent of the treatment group reporting there should be more mixing, 29.9 per cent reporting people mixed enough and 44.8 per cent reporting there should be less mixing. Respective rates of reporting for the control group were 25.2 per cent, 31.0 per cent and 44.3 per cent.

Overall, participating in the intervention appears to have a positive impact on the level of trust among one's own family and local community. This latter finding is promising and specifically relevant to one of the key aims of the intervention, which was to increase participation in the local community; increasing trust in the

community may contribute to fulfilling this aim. No differences were detected in relation to trust among people from different backgrounds, however this may be explained by the fact that trial participants were relatively homogenous. As such the intervention offered limited opportunity for additional contact with people from different religions/countries.

4.2.4.6 Participation and belonging
The survey explored participants' levels of belonging both locally and nationally. No differences were found between the treatment and control groups on either measure with very high rates of belonging reported (with regards to feeling part of the local area, 97.4 per cent of the treatment and 93.5 per cent of the control reported doing so; Fisher's p=.170, 95% CI [-0.01, 0.09]; for national belonging 97.4 per cent of treatment and 96.8 per cent of control group participants reported likewise; Fisher's p=1.00, 95% CI [-0.04, 0.04]). Such high rates of belonging limit the scope to achieve any further measurable impact on these measures. The reason for these levels of belonging are far from clear, though may be influenced by the participants answering the relevant questions in a manner which they feel they are expected to. Equally, it may also be reflective of the motivations of the participants to want to learn English. That is to say, participants already have a strong sense of belonging to their local community and the UK and therefore feel that they should learn to speak English to reflect this. In any case, there is certainly a case for further qualitative exploration of participants' and other similar populations' perceptions of identify and belonging.

4.2.4.7 Skills, employment and training
Finally, the survey explored participants' interest in attending future courses and employment intentions. There was considerable interest across both groups in all training opportunities and no significant differences were detected between the treatment and control groups in relation to future training or employment intentions.

The overwhelming majority of participants wanted to participate in all training mentioned. 84.3 per cent of the treatment group were interested in employment and skills training compared to 79.1 per cent of the control group (χ^2_1=1.08, p=.299, 95% CI [-0.14, 0.04]). Likewise, 79.9 per cent of the treatment group were interested in paid employment compared to 80.9 per cent of the control group (χ^2_1=0.01, p=.910, 95% CI [-0.08, 0.11]). 72.9 per cent versus 80.2 per cent wanted to volunteer (χ^2_1=1.96, p=.162, 95% CI [-0.03, 0.17]) and 87.0 per cent versus 84.1 per cent were interested in digital training; (χ^2_1=0.33, p=.565, 95% CI [-0.11, 0.05]).

It is perhaps unsurprising that there was little difference between the groups given that all participants were motivated individuals who had registered for English language training and actively participated in the trial.

5 Discussion

This study tested whether an intensive 11-week Community-Based English Language (CBEL) intervention was effective in supporting individuals from communities with very low levels of functional English to improve their ability to communicate in English and integrate into their wider community and society. Overall, the intervention achieved success across both of these measures. This section discusses findings within each measure in more detail. It also presents the limitations that this study is subject to and guidance on the interpretation of findings.

5.1 Impact on English proficiency

The evaluation found a strong and sizeable difference between the treatment and control groups both in terms of change in proficiency and overall proficiency at the end of the trial across all measures (speaking and listening, reading and writing). Accordingly, the analysis presented suggests the rejection of the null hypothesis with findings showing a clear impact:

- attending the CBEL intervention increased the level of change in English language proficiency (accepting the alternative hypothesis H_{1a});

- attending the CBEL intervention improved the observed level of English language proficiency (accepting the alternative hypothesis H_{1b}).

The measure of reading proficiency was subject to an unusual distribution of scores and had a statistically significant difference at baseline between the treatment and control groups. Reasons for this were not clear; however, the lack of differences in the other proficiency measures suggests that the difference was not attributable to the randomisation.

As such while differences observed with regards to reading scores alone may not reliably assess the impact of the intervention, the statistical models fitted to the reading data, which controlled for differences at baseline, showed treatment group participants made greater progress in reading and had higher scores at follow-up compared to the control group. This therefore allows us to confidently infer that the intervention positively impacted on both improvement in and overall reading proficiency.

While changes in English proficiency between the baseline and follow-up measures were significantly larger for the treatment group, a positive and statistically significant change was also detected for the control group. Such improvement may in part occur naturally over time, but another explanation for change within the period is a result of motivation associated with being a part of the trial itself. As discussed in the accompanying process evaluation there was little evidence of control group contamination insofar that the control group participants did not have access to course materials nor were they able to attend any classes during the intervention

period. [24] Control group participants did, however, appear to be developing their English proficiency outside of the intervention, often reporting that they were preparing for the start of the course in September 2016. Preparatory actions reported by control group participants included, using English more often within the household (particularly when talking to children), helping school aged children with homework, and watching more English language television, all of which is likely to have had an impact on their English proficiency. It is difficult to control for such responses within a study of this type. However, the critical fact remains that, despite the progress made by the control group, the improvement within the treatment group was still significantly greater in magnitude.

Analysis showed that the baseline English proficiency score was also a strong predictor of the level of change in scores between measures and at follow-up (across all types of proficiency). Notably, participants who had high baseline scores achieved smaller gains at the point of the follow-up. This could be due to a number of factors. Firstly, all assessment measures used had a maximum ceiling of 9, which meant those scoring highly at baseline had less scope to improve their score compared with those who scored lower. Secondly, it could reflect that this particular intervention was most effective for the target population; specifically, individuals with pre-entry level English proficiency.

Having school-aged children (aged 5 and over) was a significant predictor of improvement in proficiency, which may be indicative of the likely increased exposure to English language within the home with their child or through interaction with the child's school. Higher educational attainment was also a significant predictor of improvement in proficiency which could be indicative of a familiarity with class based study or aptitude to learn.

5.2 Impact on social integration

A range of measures were included in the survey to measure social integration. Findings indicate that there were strong and significant differences between the treatment and control groups across a number of the social integration measures. These included improvements in social interactions and everyday activity, increased confidence in interacting with health services, and higher levels of trust in the local community and among people from a different background. The analysis therefore supports the rejection of the null hypothesis and acceptance that there is an observable difference in the levels of social integration or activity through participation on a Community-Based English Language intervention, albeit not against all measures of social integration.

Analysis of the social integration measures suggests that there was a significant impact on the level of social interaction involving speaking and listening in English, and with people from different backgrounds, including evidence of more social bonds being formed. These findings may have been influenced by the impact of the intervention on social interactions (although questions specifically referred to

---

[24] https://www.gov.uk/government/publications/community-based-english-language-programme-a-randomised-controlled-trial--2

conversations held outside the classroom, participants' experiences in the classroom may have affected their responses). While this may inflate some measures, questions regarding the development of friendships and social bonds do not carry the same caveats. Indeed, as a Community-Based activity aimed at increasing social mixing, whether this occurred inside or outside the classroom is less relevant than the finding that the intervention is an effective lever for forming wider social bonds.

Analysis also showed that involvement in a Community-Based English Language intervention improved confidence in engaging with health professionals. Treatment group participants involved in the accompanying process evaluation research provided tangible examples of interacting with such actors, something that they suggest they would not have done prior to the intervention. Examples provided included making appointments with a GP and a dentist independently, and, in another example, building up the confidence to go and join a gym.

Finally, involvement in the intervention also resulted in a significant increase in the level of trust in the local community and family, though this did not transfer to trust in people from a different background or culture. It is unclear why this may be, though one cannot rule out the homogeneity of trial participants as a result of the explicit and indirect eligibility criteria used for this trial.

5.3 Limitations

There were some limitations to this study, which have previously been discussed, which may have influenced findings of this study.

First, in order to make this trial and intervention feasible the study employed a restricted randomisation process, whereby individual trial participants were randomised as part of groups of known participant 'clusters' and also blocked within delivery centres. The analytical approach used multilevel models to account for the intra-cluster correlation that may occur both within the participant groups and the delivery centres. This moderated the influence of the restricted randomisation process on the research findings.

Secondly, the process of randomisation was distinct from the collection of baseline data (randomisation took place before the collection of baseline measures). As a consequence, baseline measures were not collected from all trial participants making it impossible to account for attrition bias between the point of registration, randomisation and when the baseline measures were obtained. Overall, 18.4 per cent of trial participants who registered on to the Talk English Together course failed to provide baseline measures. From data collected at the point of registration there appeared to be only nominal differences in the socio-demographic profile of trial participants between these points; however it is not possible to make a similar assessment of English proficiency or social integration.

Related to this, while blinding the delivery provider to participant condition assignment would not have been possible in this type of intervention, the disjoint between randomisation and baseline meant that it was also not possible to blind the independent English language assessors. It is therefore not possible to exclude the

possibility that the lack of blinding at the baseline stage influenced the baseline measures.

The analytical approach adopted, specifically the introduction of baseline scores into the statistical modelling, controlled for any variance in score at the baseline stage improving the reliability of the findings presented. It is however, recommended that future studies of this nature should consider whether it is possible to allow for baseline measures to be collected at point of entry into the trial and permit blinding of condition assignment from the independent assessors.

Thirdly, the trial employed bespoke measures of English proficiency which assessed ability from Pre-entry to Entry level 2. It was not possible for participants to achieve a score of greater than 9 (indicative of 'established skills' at Entry level 2). A strong negative relationship was identified between baseline English proficiency scores and the change in score achieved at the follow-up stage. This is likely (at least in part) to be a result of the maximum ceiling that trial participants could achieve. As such, the assessment tool itself is likely to have suppressed the true level of proficiency assessed. Further development of the tool (i.e. expansion to include assessment to at least Entry level 3) is recommended to enable assessment above this parameter.

Finally, the time-bound nature of the trial, which was constrained to the duration of the course, was unlikely to be sufficiently long to capture the full impacts of the intervention. In particular, changes in social integration may further manifest over a longer period, while the trajectory of improvement in English proficiency may alter following the completion of the course. Follow up work over a longer period is therefore recommended.

5.4 Generalisability & interpretation

This evaluation has demonstrated a strong and clear positive impact that attendance on an intensive 11-week Community-Based English Language course has on both English proficiency and social integration for those with relatively low levels of English proficiency.

Though trial participants were from specific target groups (predominantly female, of Indian sub-continent descent, and of Muslim faith) findings with regards to English language proficiency are likely to be applicable to other population groups with relatively low levels of English. The same may also be true for social integration measures, though it must be acknowledged that such improvements involve more complex personal and social influences and therefore may not be as apparent among other populations. However, while the intervention was primarily focussed on improving English proficiency, it did provide opportunities for greater social interaction and stimulated community activity. Therefore, it could be expected that some improvements in integration would be made in other populations with low levels of English proficiency. Given the importance of increasing the English language ability of migrants and members of ethnic minority communities, these results are encouraging for future programmes of this nature.

Bibliography

Carvajal, SC. Baumler, E. Harrist, RB. Parcel, GS. 2001. *Multilevel models and unbiased tests for group based interventions: examples from the Safer Choices Study.* Multivariate Behavioural Research 36(2).

DCLG. 2012. *Creating the Conditions for Integration* London: HM Government

Goldstein, H. 2011. *Multilevel statistical models (4th edition)* Chichester: Wiley

Rasbash J, Charlton C, Browne WJ, Healey M, Cameron B. 2009. *A users guide to MLWiN Version 2.1* Bristol: Centre for Multilevel Modelling.

Wears, RL. 2002. *Advanced statistics: statistical methods for analysing cluster and cluster randomised data* Academic Emergency Medicine 9(4), 330 – 341.