Education Endowment Foundation

Talk of the Town

Evaluation report and executive summary

May 2016

Independent evaluators:

Allen Thurston (Queen's University Belfast)

Queen's University Belfast

The Education Endowment Foundation (EEF) is an independent grant-making charity dedicated to breaking the link between family income and educational achievement, ensuring that children from all backgrounds can fulfil their potential and make the most of their talents.

The EEF aims to raise the attainment of children facing disadvantage by:

- identifying promising educational innovations that address the needs of disadvantaged children in primary and secondary schools in England;
- evaluating these innovations to extend and secure the evidence on what works and can be made to work at scale; and
- encouraging schools, government, charities, and others to apply evidence and adopt innovations found to be effective.

The EEF was established in 2011 by the Sutton Trust as lead charity in partnership with Impetus Trust (now part of Impetus – Private Equity Foundation) and received a founding £125m grant from the Department for Education.

Together, the EEF and Sutton Trust are the government-designated What Works Centre for improving education outcomes for school-aged children.

For more information about the EEF or this report please contact:

Danielle Mason
Head of Research and Publications
Education Endowment Foundation
9th Floor, Millbank Tower
21–24 Millbank
SW1P 4QP

p: 020 7802 1679
e: danielle.mason@eefoundation.org.uk
w: www.educationendowmentfoundation.org.uk

About the evaluator

Professor Allen Thurston, Director of Centre for Effective Education, Queen's University Belfast. Allen is a former primary school teacher who now undertakes large-scale randomised trials in education. He has held numerous research grants including recent grants from the Economic and Social Research Council and the Education Endowment Foundation. Allen was responsible for evaluation co-ordination, final selection of measurements and writing of the final evaluation report.

Dr Cary Roseth, Associate Professor of Educational Psychology, Michigan State University, and Visiting Senior Research Fellow at QUB. Cary is widely respected in the field of educational research. He has pioneered the use of randomised trials in the USA and is an expert in hierarchical linear modelling (HLM) analysis and meta-analysis. Cary designed the equation for modelling data and undertook HLM analysis of the final data set blind to condition.

Dr Liam O'Hare, Senior Research Fellow in the Centre for Effective Education & Improving Children's Lives Initiative at Queen's University Belfast. Liam has substantial experience as a principal investigator on a range of randomised controlled trials (RCTs) and cluster RCTs with associated process evaluations and is widely published in the area. He has completed nine large-scale trials to date and is currently linked to three further RCTs being conducted for the EEF. He also has substantive expertise in psychometrics, particularly in the assessment of literacy. Liam coordinated applications for data from the National Pupil Data Base and assisted in preparing the final report.

Dr Jennifer Davison and Dr Patrick Stark shared a research assistant post. Other staff from the Centre for Effective Education were involved in the project, including Dr Nicole Craig and Ms Ciara Keenan. Both Davison and Stark have extensive experience of working with young people to gather research data. Stark and Craig (who collected post-test ACE data) have extensive experience of undertaking language-based tests in school-age populations. In addition, 12 field workers who were all qualified Speech and Language Therapists (SaLTs) were employed to collect post-test ACE data. Stark handled the application to collect NPDB data and identified and collated Key Stage 1 and Key Stage 2 spelling and grammar (SPAG) tests and phonics assessment data for pre and post-test national tests. Stark devised (under the supervision of Thurston and O'Hare) and issued a postal questionnaire to 128 teachers, analysed the data from the questionnaire and prepared a report of findings. Davison undertook interviews with teachers, members of the school senior management teams and local authority representatives.

Administration and support staff were involved in supporting the evaluation. They booked accommodation and travel for staff, and processed finances and claims for reimbursement from the EEF. Niki McKnight and Joanne O'Keeffe undertook these tasks.

Contact details for correspondence:

Allen Thurston
Professor of Education
Director of Centre for Effective Education
Queen's University Belfast
20 College Green
BT7 1LN

Email: a.thurston@qub.ac.uk Tel: +44(0)28 9097 5169

Contents

About the evaluator 2
Contents 3
Executive Summary 4
Introduction 6
Method 16
Impact Evaluation 28
Process evaluation 39
Conclusion 45
References 49
Appendix A 53
Appendix B 59
Appendix C 65

Executive Summary

Talk of the Town is a community led approach to supporting the speech, language and communication (SLC) skills of children and young people, aged 0-19, living in areas of social disadvantage. It is delivered by the Communication Trust.

This evaluation reports on a randomised controlled trial undertaken with 2696 pupils in 64 primary schools across three local authorities in England. The trial took place between September 2013 and July 2015. The participating pupils were in Years 2 and 5 at the beginning of the project.

The *Talk of the Town* model is a whole school approach in which four main strands or components of work are embedded:

- workforce development for all staff to support children's SLC skills, including an experienced, trained and managed Speech and Language Therapist (SaLT) working in school one day per week training all staff;
- the early identification of children's speech, language and communication needs, including developmental delays;
- universal approaches and targeted SLC interventions to support the development of age appropriate SLC skills; and
- support for senior leaders to embed speech, language and communication as part of whole school development planning and practice, to ensure a sustainable approach to speech, language and communication support.

The primary aim of this study, funded by the Education Endowment Foundation, was to identify whether Talk of the Town had an impact on pupil reading levels. Secondary aims included assessing the impact of the intervention on the language outcomes of children with low prior attainment in literacy, and exploring evidence of potential links between speech, language and communication and children's reading attainment. The study also involved a process evaluation which involved interviews a survey with participating staff.

Key Conclusions
1. There is no evidence that *Talk of the Town* had an impact on pupil's reading comprehension.
2. There is no evidence that *Talk of the Town* had an impact on oral language skills for children identified as having weaker reading comprehension skills.
3. Teachers valued the input and resources provided by The Communication Trust.
4. Teachers reported that the targeted interventions did not always provide the right level of challenge to the selected students.
5. There is scope for further research on the fact that those with low literacy were more likely to move schools.

How secure is this finding?

Security rating awarded as part of the EEF peer review process

Overall, the findings from this evaluation are judged to be of moderate to high security. The trial was set up as a randomised controlled trial in which schools that received the intervention were compared to schools operating in 'business as usual' conditions. The negative result of the primary outcome is too small to be detected by the trial and may have occurred by chance. The trial was well designed and powered, giving confidence that the programme did not have a substantial negative or positive effect on reading comprehension. Only one school in the intervention group and one school in the control group dropped out of the trial. There was, however, greater drop out of individual students, with many unavailable for testing at the end of the intervention because they had moved schools. The final sample of pupils used for analysis was well balanced.

What impact did it have?

There was no evidence that *Talk of the Town* led to better reading levels as assessed on the New Group Reading Test (GL NGRT), a test of reading comprehension, during the time period of the intervention. Although the analysis indicated a negative effect, the effect size was very small, and the result may have occurred by chance, meaning it is not possible to say with confidence that the intervention had a negative impact on reading comprehension.

There was also no evidence that *Talk of the Town* led to better led to better spoken language skills among children identified as having weaker reading comprehension skills. Six children who had scored in the bottom half of each class on the GL NGRT at the beginning of the intervention were randomly selected to complete a secondary measure in spoken language, the Assessment of Comprehension and Expression (GL ACE). The impact of Talk of the Town on this secondary outcome for this group of pupils was small and may have occurred by chance. It is therefore not possible to say with confidence that the intervention had a positive impact on oral language for these pupils. When examining the relationship between the two, it did not appear that improvements in spoken language and comprehension of language had translated into better reading scores as measured by NGRT during the time period of the intervention.

There was evidence that teachers, members of school senior management teams, speech and language therapists and local authority managers perceived that the Talk of the Town intervention had a positive effect on pupil learning and confidence. In interviews conducted midway through the intervention, there was evidence that the some teachers felt that the individual targeted SLC interventions may not always have been optimally matched to children and that there may be a case for re-evaluating the processes that facilitated targeting the interventions and differentiating approaches to ensure challenge and progression. In the paper-based survey at the end of the intervention, 90% of respondents were happy that their school had adopted the programme.

How much does it cost?

The yearly cost per school of the programme was £13,244.46 (calculated by converting the cost over 18 months to a cost over 12 months by multiplying by 2 and dividing by 3, and dividing that cost by the number of schools which received the intervention, which was 31). The cost per student of £50.97 was obtained by dividing the cost per school by the mean number of pupils per school, which was 259.86.

Group	No. of pupils	Effect size (95% confidence interval)	Estimated months' progress	Security rating	Cost
Treatment vs. control, NGRT	2696	-0.03 (-0.46, 0.40)	-1[1]	🔒🔒🔒🔒🔓	**£**
Treatment vs. control (FSM), NGRT	300	-0.09 (-0.73, 0.55)	-1	-	-
Treatment vs. control, ACE	292	0.11 (-1.19 1.41)	2	-	-
Treatment vs. control (FSM), ACE	125	-0.23 (-2.27, 1.81)	-3	-	-

[1] *Since this report was published, the conversion from effect size into months of additional progress has been slightly revised. If this result was reported using the new conversion, it would be reported as 0 months of additional progress rather than -1. See here for more details.*

Introduction

Overview of language and literacy standards in the national curriculum

The national curriculum in England specifies that the overarching aim for English is to promote high standards of language and literacy. This should be achieved by equipping pupils with a strong command of both spoken and written English, and by fostering a love of literature and reading for enjoyment. The curriculum outlines standards for language and literacy (reading, writing, spoken language) across Key Stages. During KS1, Year 1 pupils should be taught to sound and blend unfamiliar printed words and new grapheme-phoneme correspondences, and begin to understand "common exception words". Pupils' writing is expected to develop at a slower pace during this year, while they build knowledge of word sounds and the physical skills for handwriting. During Year 2, pupils should develop accurate and speedy word reading skills, and should be able to listen to and discuss a wide range of stories, plays etc. Pupils should be able to compose sentences orally and write them down, and make phonologically plausible attempts to spell unfamiliar words. During lower Key Stage 2 (Years 3 and 4), pupils should be able to read books at an age-appropriate level, and should be able to decode most new words outside of their spoken vocabulary. By the end of Year 4, they should have developed understanding and enjoyment of stories, poetry and plays, and be able to independently justify their views about what they have read. They should be able to write down their ideas with accuracy and good punctuation. During upper Key Stage 2 (Years 5 and 6), pupils should be able to read aloud a wider range of poetry and books at an age-appropriate level, with accuracy and at a reasonable pace. Pupils should be able to read with appropriate intonation to show their understanding, and should be able to summarise a familiar story in their own words. Pupils should be able to write down ideas quickly. Teachers should emphasise pupils' enjoyment and understanding of language, especially vocabulary, to support their reading and writing. Throughout these stages, the curriculum specifies that pupils who struggle to gain the necessary decoding skills should be taught urgently through a rigorous and systematic phonics programme.

Within the area of spoken language, during Years 1 to 6, the curriculum specifies statutory requirements for how pupils should use relevant questions and strategies to build their understanding, knowledge and vocabulary, and to respond appropriately to adults and peers. Pupils should be taught to give well-structured descriptions, explanations and narratives and to speak audibly and fluently, while maintaining attention during conversations, discussions, presentations etc. They should also be taught to evaluate different viewpoints and select appropriate registers for effective communication.

Progress of students is recorded in independently marked national tests at the end of KS2. In 2014, 20% of students nationally failed to reach the expected standard of 'Level 4' in grammar, punctuation and spelling, 13% failed to make the grade in writing and 11% failed to reach the grade in reading (Department for Education, 2015). There is therefore a need to find alternative ways of raising literacy attainment for those failing to reach expected standards.

Intervention

Talk of the Town

Talk of the Town (TotT) is an integrated approach to supporting the speech, language and communication skills of all children and young people, aged 0–19, living in areas of social disadvantage. It has four core aims;

- Early identification of children with speech, language and communication (SLC) needs

- A joined-up approach between parents and practitioners across health and education
- Improved outcomes for children and young people, with improved SLC skills
- A sustainable approach, so that SLC are integral to long-term planning and practice.

TotT was initially piloted across a federation of schools (one secondary, three primaries and two nurseries) in an area of social deprivation in Wythenshawe, South Manchester from April 2011 until July 2012. The pilot study was one of three strategic projects delivered by The Communication Trust during *Hello*, the national year of communication funded by the Department for Education.

The pilot study provision comprised support for all phases, from the early years through to KS4. There was also a strong emphasis on working with parents and a local community focus to engage them in support for speech, language and communication. An independent academic evaluation of the TotT pilot study, completed by Manchester University, indicated positive outcomes against all four of the core project aims.

The TotT approach was then tested in a research project funded by the Education Endowment Foundation (EEF) to explore evidence of potential links between speech, language and communication and children's reading attainment, and to look at the impact of the TotT approach on the language, of primary school children aged 5–11.

The RCT did not include a strong focus on working with parents, as in the pilot study. Specific interventions and approaches were also chosen for the randomised controlled trial (RCT) that had a moderate level of evidence as indicated by the *What Works* database; this was to ensure there was comparability within this part of the approach.

The *Talk of the Town* approach has four main strands or components of work:

1. Workforce development for all staff to support children's SLC skills. This includes an experienced, trained and managed Speech and Language Therapist (SaLT) working in school one day per week in training all staff at a universal level, providing coaching and mentoring to embed universal best practice, and supporting staff to implement targeted interventions for children with language delay.
2. The early identification of children's SLC needs, including developmental delays. This includes ongoing training and coaching support for all staff, and a range of techniques, resources and assessment tools such as The Communication Trust Progression Tools which were key in supporting identification and tracking of SLC skills.
3. Universal approaches and targeted interventions in schools to support the development of age-appropriate SLC skills, including teaching resources for teachers and teaching assistants (TAs) and evidence-informed universal approaches and targeted interventions with at least a moderate level of evidence as indicated by the *What Works* database.
4. Support for senior leaders to embed SLC as part of whole-school development planning and practice, to ensure a sustainable approach to SLC support. This includes access to termly cluster meetings with the project team as well as ongoing support from the SaLT, to help embed TotT in all classrooms and within the overall culture of the school.

To facilitate the development of these four strands as part of TotT, each school was provided with the following:

- Access to a speech and language therapist (SaLT) provided for one day a week in each school. The SaLTs were commissioned from the local NHS service to support the TotT RCT. The role of the SaLT was different from that of a 'traditional' SaLT in that they were not working to support individual children with clinical language needs, but instead worked at a universal and targeted level of input to:
  - Provide staff training through half-termly twilight meetings for all staff
  - Observe classroom practice and provide feedback to teachers with suggestions for further integrating children's SLC skills
  - Plan classroom activities with teachers to integrate children's SLC skills
  - Model strategies to promote children's SLC skills
  - Enable collaborative working between TAs and teachers to support universal best practice
  - Support staff in selecting universal strategies and targeted interventions
  - Help staff to maintain fidelity to targeted interventions
  - Support staff in identifying children with SLCN and in extending their knowledge around SLC / typical language development
  - Support collaborative practice with parents.
- An initial 'context analysis' to establish a baseline for the school, from which to track and measure their progress with implementing TotT. The context analysis comprised a range of assessments, observations and analysis to look at current school approaches and strategies for supporting SLC, to identify staff confidence in supporting and identifying SLCN, and to look at the communication environment of all classes. These activities were undertaken in partnership between the school, the school coordinator for TotT and the SaLT.
- A selection of targeted interventions and universal support designed to build on current practice.

Universal provision

TotT provided a universal programme for all staff, designed to focus on the ability of staff to provide increased support for SLC within the educational setting. This was achieved through close collaboration between the SaLT and staff, to identify and develop areas of focus over the progression of the project.

The universal provision aimed to build on what was already happening in school, as identified within the initial context analysis. The universal component of TotT was designed to ensure all children (including those with identified SLCN) had appropriate language learning environments, interactions and opportunities. It was also designed to develop staff confidence and competence in identifying children with SLCN and successfully help the teacher/TA to support the communication development of all children in class.

The universal provision included the following;

- Structured classroom observations, to help the teacher/TA reflect on the communication and language learning environment. Classroom observations were conducted by the SaLT (sometimes in conjunction with a representative from senior leadership) at periodic intervals through the project, using the Communication Supportive Classroom Observation Tool—an evidenced audit tool. The observations formed the basis of a communication-focused development plan for each class teacher. The therapists worked with all teachers to set communication targets, and to review these. This included offering suggestions of practical strategies to integrate support for children's SLC skills in class.

- A series of staff training sessions ran in every school at the start of each half term across the project (ten twilight training sessions in total). The training sessions were delivered by the SaLT and were designed for all staff to attend. Each training session focused on a different universal topic (e.g. supporting children's listening, teaching vocabulary, supporting talk in the classroom). The aim of the training sessions was to introduce an area of language, to explain why it was important and to provide staff with practical strategies, approaches and resources for the classroom to embed support for communication across the curriculum. The SaLT worked with all class teachers following the training sessions, providing in-class mentoring and support as needed, to help get the strategies and resources up and running in every class to support all children's communication development.
- The universal programme for each school included an explicit focus on training staff in early identification so they could identify children who were struggling with their SLC development. Support was provided to all staff to help them develop their knowledge and understanding of typical SLC development, to help them recognise when a child was not developing SLC skills in line with normative expectations and to ensure appropriate support was put in place, e.g. accessing a targeted intervention programme, or making an onwards referral to the specialist SaLT service. The early identification programme included a universal staff training session in the first year of the project, alongside a range of practical resources and assessment tools which the SaLT supported all staff to use. An integral element of the early identification package was the Progression Tools which aim to provide a relatively quick way for staff to determine where children are at with their SLC development in relation to age-related expectations. Staff were supported by the SaLT to use the Progression Tools to profile the SLC skills of children whose SLC development might be causing them concern.
- Support was provided by the SaLT to Senior Leaders, as well as to the nominated TotT Communication Lead, to look at ways of incorporating provision for SLC within whole-school planning and policy. Termly cluster meetings with the TotT project team took place for Senior Leaders and Communication Leads, in order to support project fidelity, sharing of feedback, shaping of the project and to ensure that any issues were identified early and managed collaboratively. The meetings also provided an opportunity for sustainability planning. Alongside this, Senior Leaders and the nominated school Communication Lead(s) were encouraged to complete regular Learning Walks to look at provision for SLC across the school.

Targeted interventions

The four targeted intervention components were these:

1. Talking Time—a targeted intervention for children in the Early Years Foundation Stage with language delay. Speech and Language Therapists worked predominantly with TAs to deliver the intervention sessions and activities. The intervention comprised two separate components: Group Talk and Story Talk. Both Group Talk and Story Talk were delivered in sessions lasting 15 minutes, at least three times a week for up to ten weeks. Talking Time focused on the development of spoken vocabulary, the ability to make inferences and developing the ability to recount a narrative. Activities encouraged the development of use of nouns and verbs. The SaLT and TAs modelled language use through the use of careful questioning and feedback for students. There was also a focus on instruction giving and receiving, and use of story without pictures to strengthen student use of inference and ability to hypothesise about stories. All of the books in the Talking Time intervention had lots of pictures, to enable the children to comment on what they could see, rather than the session requiring them to read the words. The plan of this was to strengthen inference and hypothesis. The sessions were

delivered to small groups of children (about five per group). The sessions were initially modelled by the SaLT. After training, the TA took ownership of the delivery of Talking Time.

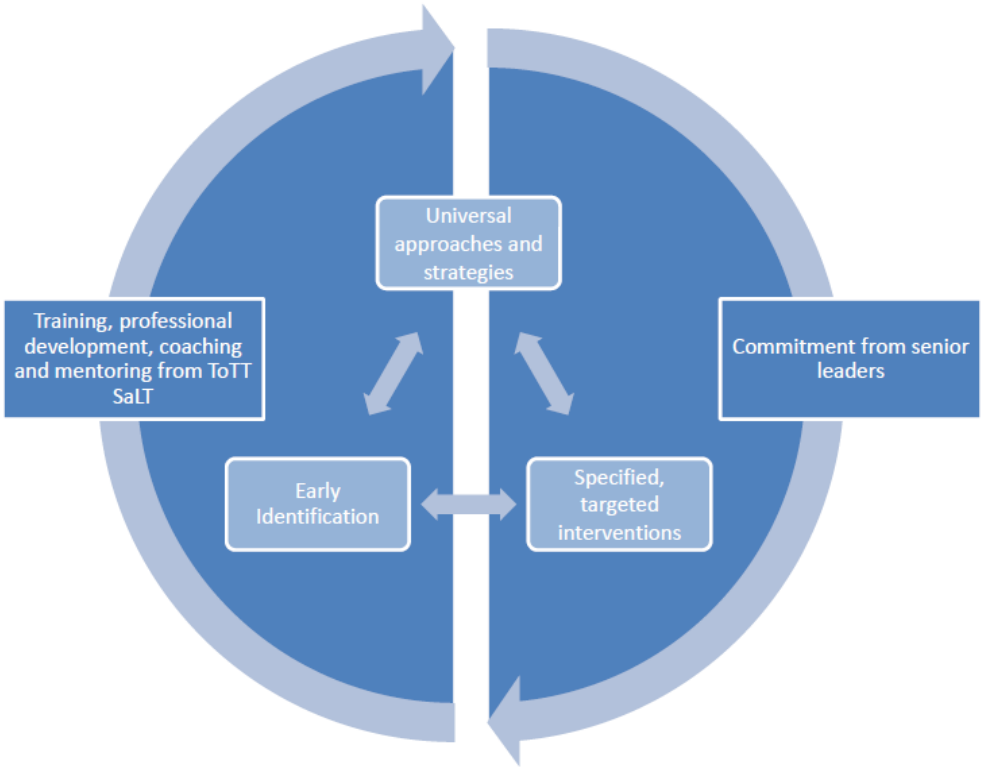
2. Talk Boost—a targeted intervention for children in KS1 aged 4–7, comprising30 sessions that took place three times per week, over a ten-week period. Each session lasted 20 minutes with a maximum of four students in each group. Sessions were targeted at children who had been identified as having delayed language skills, and targeted five different aspects of language: Attention and Listening, Vocabulary, Building Sentences, Telling Stories, and Conversations. Teachers and TAs took part in a one-day training session with the SaLT. Following training, sessions were TA-led with support from the SaLT if requested.
3. KS2 spoken language intervention—based on the York Reading for Meaning research and targeted at children in KS2. It comprised 30 sessions, delivered three times a week over a ten-week period. The intervention was initially targeted for upper KS2 (Years 5 and 6) with an adapted version provided later to differentiate some of the activities for children in lower KS2 (Years 3 and 4). The programme was aimed at children with delayed language skills and targeted the following aspects of spoken language: Vocabulary, Talking and Understanding, Figurative Language, and Spoken Narrative. The intervention was based around a text. Teachers and TAs attended an initial two-hour twilight training session, led by the SaLT, to support them with running the intervention. Following training, TAs led the targeted intervention with input from the SaLT as required.
4. Phonological Awareness—targeted at all stages. This was a flexible programme of work to support phonological awareness at different levels dependent on age and ability; the intervention comprised 15 weeks, with a minimum requirement to complete 10 consecutive weeks. The intervention was delivered three times a week, with each session lasting no more than 30 minutes. It was aimed at KS1 or KS2 pupils for whom it may have been developmentally appropriate. The first session of the week was modelled by the SaLT, with the TA delivering the second and third sessions.

Children were selected for the intervention groups through collaboration between the SaLT and the teacher/TA. For each targeted intervention, inclusion criteria were provided to support staff in selecting the children who would benefit most from the groups. With the exception of the Phonological Awareness intervention, the targeted intervention programmes were designed to support those children with delayed language to improve their targeted SLC skills, with the aim of developing age-appropriate skills in these areas. Progression Tools were used before and after the intervention to ensure that the right children received the right intervention, and also to track their SLC progress following targeted intervention.

All of the targeted interventions encouraged whole-class links to take place, so that the class teacher reinforced in universal teaching the skills that the children had learned in the groups. The reinforcing of universal aspects of the targeted interventions adapted for the class was a really important part of the approach, to ensure that there was a link between the universal and the targeted elements of TotT. Figure 1 represents how these components interconnected.

Figure 1: Components of the TotT model and approach

Talk of the Town
Components of the ToTT model and approach



Background evidence

Speech, language and communication needs (SLCN) can significantly impact children's learning and development. SLCN prevalence varies between 1% and 15% in a typical population (Boyle et al., 1996), but this may rise to up to 50% in disadvantaged areas (National Equity Panel, 2010). In a report on the development of 11,000 children in the UK, children with poor vocabulary skills at 5 years old were four times more likely than typically developing children to be poor readers in adulthood (Law et al., 2009). Furthermore, SLCN may be more common in children who have emotional and behavioural difficulties (EBD). One review found that between 56% and 71% of children with EBD also experienced language difficulties (Benner et al., 2002). Strong behavioural needs may also mask SLCN, and children with EBD have been found to make significant progress in terms of self-esteem and behaviour (in addition to language) when provided with a language intervention (Law & Sivyer, 2003). In particular, there is concern that delays in development in speech and language are more prevalent among those from lower socio-economic backgrounds.

Measures of socio-economic status (parental income, employment etc.) are significantly associated with children's productive vocabulary from as early as 7 months to 3 years old, which in turn is predictive of later verbal ability in school (Walker et al., 1994). A study which followed children from high, mid and low socio-economic status (SES) backgrounds (Hart & Risley, 1995; Hart & Risley, 1999) found that children from low SES backgrounds had significantly slower vocabulary development, which itself was associated with lower IQ at 3 years old. These children also suffered from poorer educational achievement at ten years old. Children with low SES backgrounds in the UK are up to twice as likely to experience delayed

expressive language abilities than children from mid to high SES backgrounds (Peers et al., 1999). When looking at moderate to severe expressive language delay in children, Peers et al. found that low SES backgrounds was associated with a five times increased risk when compared with mid to high SES backgrounds.

Speech and language impairments are reported to have a wide range of negative impacts on children's development. Bishop and Adams (1990) followed the early years development of children with impaired language skills between the ages of 3 and 4 years. Children whose language had developed to a typical level by the age of 5½ years suffered no further impairments. Children whose verbal impairment persisted at 5½ years of age experienced further problems with both reading and oral skills, and reading comprehension especially suffered. Nathan et al. (2004) found that children with a history of speech and language impairments performed worse than typically developing children in their Statutory Assessment Tests (SATs) in reading, spelling and maths, at the end of Year 2. There is evidence that children whose speech and language impairments had been remediated by this time performed equally to controls with no history of impairments.

Persistent speech and language problems are also strongly associated with lower educational attainment. Snowling et al. (2001) followed a group of children who had a history of speech and language impairment since pre-school. Children with impaired language at 5½ years went on to later perform worse than controls in their GCSE examinations. In the Manchester Language Study (Conti-Ramsden et al., 2001), the educational performance of 120 adolescents with language impairments was studied throughout secondary education. Almost 25% of language-impaired adolescents did not achieve a single GCSE at the end of their compulsory education, compared to only 1% of the school population overall (The Poverty Site, 2015).

Previous reviews of speech and language interventions

In their 'What Works?' report, Law et al. (2012) found that out of 57 SLCN interventions, 3 had a strong level of evidence, 32 had moderate evidence, and 22 had indicative evidence of improved language outcomes. They concluded that the small number of intervention studies which found strong evidence may be due to their small size and the lack of large-scale effectiveness trials. Therefore, evidence as to whether students with SLCN can experience significant improvements in language development from a broad range of speech and language interventions is mixed.

Talk of the Town was previously reported to improve the identification of children who were not achieving the levels of required language in both nursery (by 35%) and primary (by 26%) in a pilot study. In addition, accelerated KS1 language scores indicating 12–14 months of age standardised progress in a 10-week period were reported for Talk Boost (Ainscow et al., 2012). During pre-planning for this evaluation raw data from these studies were examined from a data set drawn from three pilot classrooms. On the GL ACE 6–11 Test, mean *ES*s in the 'Naming section' were 0.51. On the 'Sentence comprehension' section of this test *ES*s were between 0.35 and 0.82 with a mean *ES* of 0.59. On the 'BUS Story' test (a standardised test of spoken narrative recall) *ES* were observed of between 0.15 and 0.23 with a mean *ES* of 0.19. Talking Time has been reported to improve talking of nursery age children when compared to results from children who have not undertaken Talking Time. Results indicated significant gains in these areas and showed particular improvement in talking. This study noted that the programme needed to be finely tuned to the correct level for children in order to maximise any gains that may accrue (Dockrell & King, 2007). However, neither of these studies had a control group and effect sizes (*ES*) were calculated on the basis of standardised comparison scores. Results from previous pilot studies were therefore not conclusive and interpretations of causation would be limited.

Teacher confidence in providing speech and language interventions to children

Despite their enthusiasm for involvement in speech and language interventions, teachers in the UK have reported receiving a lack of specific training to do so (Dockrell & Lindsay, 2001). Sadler (2005) interviewed 89 teachers in the North East of England (Reception, Year 1 and Year 2). Ninety per cent of the teachers reported that they had not received training for working with children with speech and language problems. Many of them (57%) felt there were restricted opportunities in their current class for working with children who had a speech and language impairment, and 63% reported that they were not confident in working with these children. Considering that collaborative efforts between SaLTs and teachers have been shown to be more effective than purely clinical treatment (Throneburg et al., 2000), it is possible that, in addition to the involvement of a SaLT, providing extra support to teachers may be beneficial to improving language outcomes in children with SLCN. Within this it is also important to note that there is an apparent need to explore the support required by TAs to enable them to work effectively with SaLTs.

Evaluation objectives

Research questions

1. What are the effects of the *Talk of the Town* approach on student reading levels in Granada Learning (GL) New Group Reading Test (NGRT)?
2. Secondary research questions:

    a. What are the effects of the *Talk of the Town* intervention on student oral language levels as measured by GL Assessment of Comprehension and Expression 6–11 (ACE)?

    b. To what extent do language levels at pre-test predict reading outcomes at post-test?

    c. How do teachers, members of school senior management teams, speech and language therapists and local authority managers perceive that delivery of the *Talk of the Town* approach can be optimised in future full-scale initiatives?

Project team

A team from The Communication Trust (CT) delivered the project. The Communication Trust describes itself as '… a coalition of over 50 leading voluntary sector organisations, which raises awareness, influences policy, promotes best practice among the children's workforce and commissions work from its members. The Trust works collectively to ensure that every child and young person is enabled to communicate to the very best of their ability. The Trust aims to achieve this by working in partnership to build skills, knowledge and confidence in the children's workforce, based on a foundation of good practice, robust evidence and effective policy and legislation.'

The project was led and managed for the CT by Wendy Lee (Professional Director), Eve Wagg (Programme Manager) and Kerry Davis (Project Officer). This management team led a wider team at CT who developed appropriate resources for schools and also the team of nine SaLTs who delivered TotT in the intervention schools. Their job was to support their assigned schools (this ranged between two and five schools for each therapist) to deliver the TotT approach as previously described.

Evaluation team

Professor Allen Thurston, Director of Centre for Effective Education, Queen's University Belfast. Allen is a former primary school teacher who now undertakes large-scale randomised trials in education. He has held numerous research grants including recent grants from the ESRC and the EEF. Allen was responsible for evaluation co-ordination, final selection of measurements and writing of the final evaluation report.

Dr Cary Roseth, Associate Professor of Educational Psychology, Michigan State University and Visiting Senior Research Fellow at QUB. Cary is widely respected in the field of educational research. He has pioneered the use of randomised trials in the USA and is an expert in HLM analysis and meta-analysis. Cary designed the equation for modelling data and undertook HLM analysis of the final data set blind to condition.

Dr Liam O'Hare, Senior Research Fellow in the Centre for Effective Education & Improving Children's Lives Initiative at Queen's University Belfast. Liam has substantial experience as a principal investigator on a range of RCTs and cluster RCTs with associated process evaluations and is widely published in the area. He has completed nine large-scale trials to date and is currently linked to three further RCTs being conducted for the EEF. He also has substantive expertise in psychometrics, particularly in the assessment of literacy. Liam coordinated applications for data from the National Pupil Data Base and assisted in preparing the final report.

Dr Jennifer Davison and Dr Patrick Stark shared a research assistant post. Other staffs from the Centre for Effective Education were involved in the project, including Dr Nicole Craig and Ms Ciara Keenan. Both Davison and Stark have extensive experience of working with young people to gather research data. Stark and Craig (who collected post-test ACE data) have extensive experience of undertaking language-based tests in school-age populations. In addition, 12 field workers who were all qualified SaLTs were employed to collect post-test ACE data. Stark handled the application to collect NPDB data and identified and collated KS1 and KS2 SPAG tests and phonics assessment data for pre and post-test national tests. Stark devised (under the supervision of Thurston and O'Hare) and issued a postal questionnaire to 128 teachers, analysed the data from the questionnaire and prepared a report of findings. Davison undertook interviews with teachers, members of the school senior management teams and local authority representatives.

Administration and support staff were involved in supporting the evaluation. They booked accommodation and travel for staff and processed finances and claims for reimbursement from the EEF. Niki McKnight and Joanne O'Keeffe undertook these tasks.

Ethical review

The School of Education Ethics Committee at Queen's University Belfast gave ethical approval for the research project. Both Research Governance and the School of Education Ethics Committee at Queen's University Belfast also approved any minor changes in procedures.

All research was conducted according to QUB School of Education Ethics Committee Guidelines. Ethical consent was obtained for the trial. Informed consent was obtained from participants using 'opt-out' procedures. A letter regarding literacy testing and access to National Pupil Data Base (NPDB) testing was sent home asking parents if they wished to opt out of the testing and/or access to NPDB access. No parents opted out of undertaking GL NGRT tests and eight parents requested that we did not access their child's NPDB records (these wishes were complied with). ACE testing was completed with opt-in consent. Some schools requested

that they give ethical permission as *in parental locus.* The Ethics Committee approved this request and some consent was obtained this way (in 17 schools).

Method

Trial design

While *Talk of the Town* is a whole-school approach, a pragmatic decision was taken to focus on two year groups in all schools for the evaluation. The selected year groups were from the KS1/2 transition (Year 2 through Year 3) and from KS2 (Year 5 through Year 6). The unit of randomisation was the school. This cluster randomised design was selected to compare the relative effectiveness of the intervention and control while accounting for the clustered structure of the data, with students nested within schools. A pragmatic design was required as TotT was a complex programme comprising a universal approach that was delivered in a tailored fashion in each school. Although criteria and selection procedures were used to decide which students received targeted interventions, the complexity of implementation meant that this was a pragmatic trial to determine whether the TotT programme as a whole enhanced outcomes in reading.

There was no plan in the protocol to determine what was happening in the 'control school' condition. These schools had equal access to Educational Psychology, Speech and Language Therapy Services and Health Support within their local authority to support students with special educational needs. In addition, control schools may have been able to purchase SaLT time. However, there was no systematic attempt to collect such data as it fell outside the remit of the published protocol. Schools that remained in control condition and completed post-tests were promised access to one of two interventions, either Talk Boost for KS1 or the KS2 targeted intervention, including training for staff to run the interventions to the value of £420.

Recruitment was strong for the trial and there was no need to change the initial overall design. There were a few decisions made after the initial protocol was published. One was to use fieldworkers to undertake ACE testing and to have oversight of NGRT at post-test. This was to condense the testing period on a two-week block so as to avoid clashes with national testing and school transition preparations. Other changes were the decision to invest considerable time in developing the model used to test the intervention, but this development was done blind to condition. This was necessary as despite the large sample there were still differences (albeit not significant) in pre-testing scores and variations in demographics in each school. It was important to fully understand which variables should, and should not, be included in the final model. However, an expert in statistical modelling (Prof Cary Roseth, Michigan State University) who played no part in the delivery of the approach and also no other part in the evaluation, undertook this process. Therefore, given this separation, risk of bias from redefining the model was minimised. This process is described in detail later in the report.

Participants

Inclusion criteria for schools

The trial of *Talk of the Town* ran with an intervention and a control condition with 'business as usual'. To be assigned to condition, schools had to meet the following inclusion criteria:

- had at least 20 Year 2 students in single form entry class and 20 Year 5 students in single form entry class;
- were willing to be randomly assigned to condition at the school level;
- met criteria acceptable to the EEF, in that the majority of schools in the sample would be from a disadvantaged area (meaning the overall sample would be above national mean for free school meals which was 15% in the year of study);
- were willing to supply access to socio-economic and demographic data from students;

- had access to Educational Psychology, Speech and Language Therapy Services and Health Support (note that no schools were excluded on this basis);
- had access to KS1 data for students starting Year 5 in September 2013; and
- were willing to engage with the *Talk of the Town* intervention and implement this in at least Year 2 and Year 5.

All schools that met these criteria and expressed interest in participating in the project were randomised into the study.

Recruitment strategy and consent

Recruitment took place in three local authority areas to generate the sample. Local authorities based in the North West, North East and South of England were selected as The Communication Trust (TCT) felt that strong relationships within these local authorities and schools would lead to more effective recruitment. Local authorities were selected through an application and interview process. Recruitment partnerships were formed with selected local authorities. Recruitment was undertaken and managed by TCT. Recruitment events took place in three local authorities where schools attended to receive more information. At this point schools filled in an expression of interest form that was returned to The Communication Trust and were selected for the study using the selection criteria. QUB supported TCT in the recruitment process. QUB provided input to the experimental design and measures to be used. The recruitment events generated 70 expressions of interest. The aim was to recruit 62 schools. Two schools were rejected, as they did not fit the inclusion criteria. Two schools withdrew before selection for the study and randomisation was undertaken. Of the remaining 66 schools, two were selected as 'reserve schools' in case of drop-out at an early stage of the project and Allen Thurston at Queen's University Belfast randomised the remaining 64 schools to condition independently from The Communication Trust. The schools selected as 'reserve schools' were those who signed up last within the local authority where they were based.

Schools signed a memorandum of understanding with TCT. This detailed school responsibilities, timelines and expectations of schools, TCT and evaluators. Individual consent was not sought by QUB for being involved in TotT. This was delivered by TCT and was part of the normal schoolwork in developing the speech and language of children. It was therefore the teacher's professional choice as to whether to deliver the intervention or not. Opt-out consent to collect data from test results and from the National Pupil Data Base was sought from parents. Letters were sent home through the schools. Opt-in consent was sought to undertake ACE testing. As previously detailed in 'Ethical review', some headteachers/teachers gave this consent as *in parental locus.* Examples of consent letters are provided in Appendix A.

Outcomes

One primary and one secondary outcome measure were used in this evaluation and additional data was obtained from the National Pupil Data Base. Instruments and measures are summarised in Table 2.

- The primary outcome measure of the study was reading as measured by the Granada Learning (GL) New Group Reading Test for Year 2/3 and Year 5/6 students.
- The secondary outcome measure was language of Year 2/3 students with below average literacy as measured by GL ACE assessment.
- Tertiary outcomes measures were KS results (where available) and various subgroup analysis of the reading outcomes of the cohort by gender, ethnicity, free school meal status, school, class, year group and language level at pre-test as a predictor of the primary outcome.

Table 1: Planned instruments and measures for evaluation

Measure	Pre-test	Post-test	Related additional/ follow-up data that could be collected from NPDB
KS1 results	KS1 results from Year 5 class when they were in Year 2		KS1 results from Year 2 class at the end of the first year of implementation
Assessment of Comprehension and Expression (ACE) test	6 students per class from Year 2	6 students per class from Year 2 with below average literacy (now at the end of Year 3)	
New Group Reading Test	Year 2	Year 2 (now end of Year 3)	
New Group Reading Test	Year 5	Year 5 (now end of Year 6)	
KS2 results		Year 5 (now end of Year 6) Possible follow-up for Year 2 when they get to the end of their Year 6	KS2 results from Year 2 class (when they are at the end of Year 6) Available Jan 2016

Primary outcome: Granada Learning New Group Reading Test (NGRT)

The primary outcome measure was the age-standardised score from the New Group Reading Test (NGRT) (Burge et al., 2010). Standardised scores were selected as the primary outcome measure as opposed to raw scores. There were no significant differences in floor effects of each score (41 students having floor scores of 0 on raw NGRT test scores and 42 having the floor score of 59 on standardised NGRT test scores). However, use of age-standardised tests did have the advantage that it allowed analysis of outcomes across the whole sample using the school as a cluster for two year groups (something that would have been problematic with raw scores as standard deviations would have been expanded due to the three-year age gap between the youngest and oldest students being tested). This test of reading ability was administered at pre and post-test in all control and intervention schools with one class of Year 2 pupils in each school (post-test at the end of Year 3) and one class of Year 5 pupils in each school (post-test at the end of Year 6) students. Pre-test NGRT was implemented by teachers, but coded and processed by Granada Learning Assessment (GL) blind to condition. Post-test NGRT was implemented under the supervision of field workers and was coded by staff from GL blind to condition. Each test subject had its raw test score converted into a standardised score that had a mean of 100 and standard deviation of 15. NGRT is a test of reading comprehension that includes completing stand-alone sentences by filling in the missing word, choosing the correct word to complete sentences in the first paragraph of a passage of text and selecting the correct answer to questions about passages.

Each paper consisted of questions about what is the correct missing word from a sentence, and the test-taker must select one of five options. Afterwards, there were four comprehension

passages of increasing difficulty that were a mixture of factual and fictional texts. An advantage of using the paper version was that many primary schools would have required additional support to install and manage the online test version.

Pupils are given scores for the sentence completion and passage comprehension sections. The Rasch ability scores for the items are processed by GL, creating a scale score for the section, i.e. a 'sentence comprehension' score for the sentence completion section, and a 'passage comprehension scale' score for the comprehension passages. An 'overall scale' score is derived for the two completed sections and adjusted to allow for the pupil's chronological age to produce the Standard Age Score (SAS) compared to reference groups (not involved in the study) by Granada Learning. The SAS score was used in analysis of data from the NGRT measure. Standardisation was reported to have been conducted in 2010 (Burge et al., 2010, p.138). Standard deviation on Test 2 and 3 are identical allowing the comparisons to be made.

The pupils at pre-test were administered the paper tests NGRT 2A (Year 2) 3A (Year 5). At post-test they were administered paper tests NGRT 2B (Year 3) and 3B (Year 6).

Pre-test NGRTs were administered by the school (before they knew which condition they were assigned to). Post-test NGRTs were administered by the school under observation from a field worker employed by Queen's University Belfast blind to condition.

Tests were hand scored by employees of GL blind to condition and data uploaded onto a secure server for encrypted data transfer.

Secondary outcome: GL Assessment of Comprehension and Expression (ACE) 6–11

The secondary outcome measure was a test of language, ACE, which was undertaken with a randomly selected sub-set of six students from the bottom half of the class reading attainment profile (as measured by NGRT pre-test). Pre and post-test ACE data were collected from the six pupils in each school class in Year 2. The test was repeated with the same students at the end of Year 3 (note that it was not possible to test all students in the sample for either pre or post-test owing to the high cost of this test administration which was on an individual basis and took 45–60 minutes to administer). This measure was selected after set-up meeting discussions with the delivery team. It was noted by the delivery team that in the plan the sub-sampling of students for ACE from the bottom half of the reading attainment ranking of the class allowed the impact of *Talk of the Town* on language development to be established in those drawn from a group with lower than class mean reading attainment. The delivery team felt that this group was the most likely to benefit from TotT. However, whether these individuals were eventually selected for targeted intervention was down to their identification by the teacher/SaLT. Whether they received a targeted intervention was not tracked in any way by the evaluation team and was not noted as being required in the protocol. This was an intention to treat intervention where the treatment was the provision of an intervention that was designed to be targeted on children with delayed language. In terms of the ACE assessment the following sub-sections were assessed:

- Sentence comprehension
- Inferential comprehension
- Naming
- Syntactic formulation
- Semantic decisions

These tests were undertaken by a group of 4 field workers at pre-test and 14 field workers at post-test. Use of more field workers at post-test was to condense the testing period into a two-week timeframe in June 2015. This was to avoid clashes with national testing regimes and

school transfer for Year 6 students. At pre-test, inter-tester reliability analyses revealed alpha values of >0.98 indicating very high levels of agreement. Twelve of the post-test field workers were qualified SaLTs who were put in touch with Queen's University Belfast (QUB) by TCT. They were trained and employed by QUB. Raw test scores were calculated by testers and tables used to calculate the age-standardised score for the test. The inter-tester reliability trial for the pool of 14 field workers employed to collect post-test determined the replicability of scoring of the test and alpha values at >0.99 on all sub-scales of the ACE test were observed. Pre to post-test reliability was >0.99 when comparing reliability trials at pre and post-test for fieldworkers who collected test data at both test points. The high alpha values pre to post-test and between fieldworkers is probably a function of the structure of the test. The test comes with a manual. It is delivered with a script and answers and allowable alternative answers are prescribed in the manual. The inter-tester reliability trials were undertaken by the same person role-playing the role of a testee, giving pre-determined answers as questions were asked by a research assistant, and all those taking part in the trial coding answers for comparison by the research team. It should be noted that the use of an additional 12 fieldworkers who were recruited through TCT contacts was a change to the published protocol as noted above. However, none of these fieldworkers had been involved with *Talk of the Town* and they were employed on contracts by QUB. Their data was passed directly to QUB and they received training for the job they were undertaking at one central location. Therefore, risks of this change were minimised and managed.

Tertiary outcome: Key Stage (KS) data

The tertiary outcome measure was KS results (where available) and various subgroup analysis of the reading outcome. Mean class KS1 test scores for current Year 5 students in the quantitative measures group were used to assign schools to condition pairwise as indicated below. KS1 data was collected from students at the end of Year 2 to provide an interim report on differences between profiles in treatment and control schools. KS2 test data was collected in English and mathematics for both intervention and control groups at the end of Year 6 for students who started the project in Year 5.

Sample size

The ability to detect a treatment effect at a certain level of power depends on several factors: intraclass correlation (ρ), the proportion of post-test variance that was explained by the covariates (r), the average number of students in each school (n), and the number of school-level clusters (j).

In the power analysis used to determine sample size, an estimate of the intra-class correlation (ICC) of ρ=0.07 was used. That figure represented the proportion of variance in achievement scores that was anticipated between schools. This estimate was derived from the range.05 to .12 for low-achieving schools reported by Hedges & Hedberg (2007), and was consistent with ICCs from previous studies (e.g. Borman et al., 2007). To increase statistical power, four covariates were included. Based on data from previous studies, the proportion of variance in post scores explained by the covariates is estimated at r^2=0.60.

In addition to the main achievement analyses, the protocol stated that analyses were planned to determine interactions between treatment and pre-test, ethnicity (although this changed in the final model owing to the uniformity of the sample), gender and free school meal status. There was adequate power for each of these analyses, as the number of schools would remain identical to those for the main analyses, and power for RCTs is driven primarily by the number of school-level clusters (j) rather than the within-school students (n). It was assumed that the intervention could have an *ES* of 0.23 (taken from unpublished information provided by TCT from previous pilot studies), which then required 62 schools to participate (31 intervention, 31

control) to detect effects in reading at 80% power with alpha=0.05 and ICC=0.07 and mean class size=25 students. The sample size of 62 schools was a pragmatic sample size decision to give the best chance of detecting an effect. Two additional schools were recruited to the sample to allow for attrition, making the total sample 64 schools. Post hoc power analyses and changes to the planned model are outlined later in this report.

Randomisation

Assignment to condition was by block randomisation. Block randomisation took place pair-wise at the school level based on historical KS1 test scores for the Year 5 targeted year groups taking part in the study. Schools were rank ordered on the basis of mean KS1 scores for current Year 5 classes and assigned to condition pairwise, i.e. the top two schools were assigned one each to intervention/control, schools three and four on the list similarly assigned and so on until each school was assigned. Block randomisation was preferable as it minimised the chance of difference in gains for intervention/control groups on the basis of prior attainment. Including pre-test scores in the analysis allowed us to infer that any difference between the change in NGRT scores in the treatment group and the change in the control group can be attributed solely to the influence of the TotT intervention. Equal numbers of intervention/control condition schools with 'business as usual' schools was achieved within each local authority (by blocking schools in three groups, based on geographical area/local authority) to ensure that there was equal access to Educational Psychology, Speech and Language Therapy Services and Health Support as would have been expected in business as usual (compared to the same number of intervention schools receiving TotT).

To assign schools to condition, a random number generator application called 'The Random Number Generator' designed by Nicolas Dean (http://nicholasdeanapps.webs.com), version 3.6, was programmed to generate a number between 1 (*Talk of the Town*) and 0 (control) by Allen Thurston. It produced the number, and the school with the highest mean KS1 score in a particular local authority from 2011 was assigned to the appropriate condition. Schools were then sequentially assigned to condition using alternate assignment of schools to condition (e.g. top school KS1 2011=1, *Talk of the Town* intervention; second top school KS1 2011=0, control; third top school KS1 2011=1 *Talk of the Town* intervention; fourth top school KS1 2011=0, control…) until all pair schools had been assigned to condition. This process was repeated for lists of schools for each local authority area.

Randomisation took place before schools undertook pre-testing using the GL NGRT test (September 2013). However, schools were not informed which condition they had been assigned to until all pre-testing had taken place and tests sent for independent marking. After analyses confirmed parity between *Talk of the Town* and control conditions in historical KS1 2011 scores and current free school meal (FSM) percentage, no further manipulation of the samples was required as the intervention and control samples had no significant differences in KS1 2011 reading scores and FSM percentages.

Analysis

Multilevel modelling

The study's data can be thought of as multilevel, with students nested within schools. Accordingly, we used a series of linear mixed models (LMMs) to account for the clustered structure of the data while comparing the relative effects of the intervention and control conditions on the primary reading outcome (NGRT) and secondary language outcome (GL ACE). LMMs can be viewed as hierarchical linear models (HLMs) with individual- (level 1) and cluster-level (level 2) equations (Raudenbush & Bryk, 2002). Appendix C includes additional

information about LMMs and the specific unconditional and conditional models used for the analysis.

Under the assumption of an ignorable missing data mechanism such as MCAR or MAR, LMMs accommodate missing data using maximum likelihood-based estimation procedures. This is in keeping with EEF policy on conducting analyses on an 'intent to treat' basis. To estimate parameters and test statistics, we used the restricted maximum likelihood method (REML), and a significance level of $\alpha = .05$, in the PROC MIXED procedure of SAS 9.4 (SAS Institute, Inc., Cary, NC, USA).

One change to the model from that stated in the protocol was the decision to use a 'step-up' strategy for model building (West et al., 2014) as this would provide an empirical basis for determining how much of the variability in the outcomes was accounted for by the experimental condition and student covariates. Specifically, while blinded to the treatment condition, Dr Roseth began the analyses with an unconditional two-level model containing only a single fixed effect (the overall intercept) and random effects associated with the intercept for schools and residuals. He then tested conditional two-level models including fixed effects associated with pre-test achievement, experimental condition, and student-level covariates. He used likelihood ratio tests to decide whether to keep the fixed effects associated with the student-level covariates and interactions with the experimental condition. Residual plots showed no evidence of outliers or violations of model assumptions (e.g. normality and constant variance).

Addressing missing data

Another change from the protocol was using multiple imputation to test the sensitivity of the results to missing data. This change was undertaken because the level of missing data was greater than expected and, as detailed in the Impact section below, there was evidence of differential attrition in treatment and control, i.e. those of lower attainment in reading were more likely to be missing in both the control and intervention sample. Multiple imputation is currently regarded as the "state of the art missing data technique" (Enders, 2010, p. 343) because it improves the accuracy and power of analyses relative to other methods of handling missing data (see also Schafer & Graham, 2002). Specifically, we used the Markov chain Monte Carlo (MCMC) method (Schafer, 1997) in the SAS MI procedure to generate 50 imputed data sets separated by 300 iterations of the imputation algorithm. To account for the data's multilevel structure, we imputed data separately for each of the 64 schools. We also included students' pre-test scores, year in school, experimental condition, and five covariates (gender, free school meals, minority status, special education status, and English as additional language) in the imputation process. This is in keeping with the methodological literature on missing data (e.g. Enders, 2010; Graham, 2003) and makes the missing at random (MAR) assumption more plausible (Collins et al., 2001). After creating the complete data sets, we estimated the HLMs on each of the 50 imputed data sets and then used Rubin's (1987) formulas in the SAS MIANALYZE procedure to combine the fixed-effect parameter estimates and standard errors into a single set of results. We do not report random effects or deviance statistics for the multiple imputation data because at present the SAS MIANALYZE procedure in SAS 9.4 does not pool variance estimates.

Subgroup analysis

Subgroup analyses were conducted by testing student covariates and interactions (e.g. gender, ethnicity, etc.). In keeping with the protocol, specific decisions about subgroup analyses were based on the observed composition of the sample and the actual data analyses. For example, the protocol originally included ethnicity in the model but it was noted at pre-test that 96.3% of children fit a white or non-recorded ethnic group, which would have led to very small sub-

samples. Accordingly, the covariate, minority status, was used instead. Other student-level covariates included their individual pre-test score, year in school, gender, minority status, reading age (as measured by pre-test NGRT score), status in special education (SEN), English as an additional language (EAL), and receiving free school meals (FSM). In our view, testing interactions between the intervention and each of these subgroups is a more empirically driven approach to subgroup testing and strengthens confidence in the statistical validity results. In addition, it allowed the exploration of effects on different constituent groups of the sample. Although there were low numbers of subgroups in the sample, given the number of clusters subgroup analysis would discover significant patterns if effect sizes were large enough. Finally, we also note that the 'step-up' approach to model-building and testing for subgroup interactions was developed by Prof Cary Roseth (Michigan State University/Queen's University Belfast) independently from the rest of the evaluation team and prior to any analysis of post-test data taking, as was the stated process in the protocol. This too strengthens confidence that some student covariates were treated differently from others.

The one exception to this approach to subgroup analyses was driven by updated guidance from the EEF. Following the EEF's policy on analysis (2015, p. 3), we computed an *ES* on the primary outcome for FSM pupils using a separate model.

Also following the EEF policy on analysis (2015) and the What Works Clearinghouse (2013, p. F.9), we calculated the *ES*, Hedges' g, from the multilevel analyses by dividing the coefficient for the intervention's effect (the numerator), which represents the adjusted group mean difference adjusted for both level 1 and level 2 covariates, by the unadjusted pooled within-group SD (the denominator). Note that adjustment and not gain scores were used.

Process evaluation

The process evaluation aimed to assess whether conditions necessary (identified through a logic model drawn from documentation and discussion with TotT staff and school staff) were achieved. The logic model is presented in Figure 2. The primary and secondary measures would assess outcomes. The process evaluation was specifically designed to assess:

- whether a whole-school approach was achieved;
- whether targeted interventions were delivered;
- whether targeting of the intervention at specific students was appropriate;
- what barriers existed to achieving the required conditions for optimal implementation; and
- the opinions of teachers, members of senior management teams and local authority representatives/managers regarding how worthwhile/attractive TotT was as an intervention.

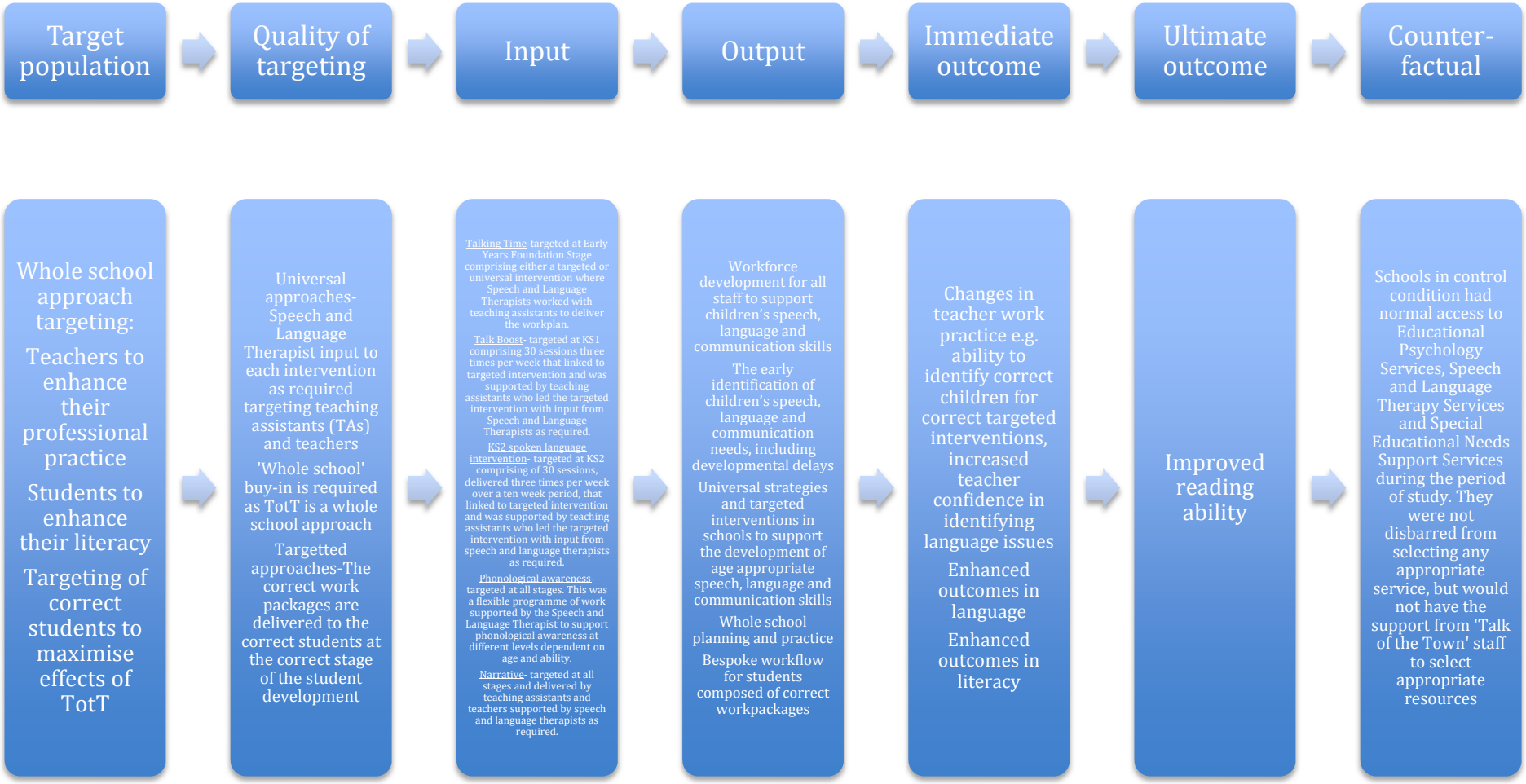
These aims were achieved with a series of interviews conducted one year into the evaluation (i.e. 12 months after the start of the intervention) and a postal questionnaire conducted at post-testing. The rationale was to collect in-depth process information from interviews during the main intervention and a more broad-based set of data at the end of the intervention.

Interviews: 20 semi-structured interviews were undertaken with Year 2, 3, 5 or 6 teachers from a class that formed part of the quantitative data set (one per school from 20 of the participating intervention schools selected randomly on a proportionate basis across each local authority) to explore their perceptions of what worked well in the project and how the intervention could be enhanced. Similar interviews were also undertaken with one member of the senior management team (SMT) from each of these 20 schools and from a senior manager from each school's local authority. The interviews were conducted by Jenny Davison (Queen's University Belfast) via telephone; they were recorded and transcribed for analysis. Jenny Davison, Patrick

Stark and Allen Thurston conducted analysis. Interviews were undertaken about 12 months into the project.

Questionnaire: A postal questionnaire was issued to all teachers who participated in the second year of the project as a class teacher in Year 3 or Year 6. Teacher perceptions of the potential benefits of the intervention were collected through a structured questionnaire. In addition, teachers were asked about the wider impact on their students. The survey was sent to 62 teachers who were delivering TotT. Responses were received from n=22, Year 3 teachers and n=21, Year 6 teachers. Seven teachers who were also TotT Communication Leads, one headteacher and one assistant headteacher also responded to the survey, giving a total of n=52 spread across 31 schools, a response rate of 84%. The postal questionnaire was designed by Liam O'Hare and Patrick Stark and analysed by Patrick Stark. The questionnaire was undertaken about 18 months into the project.

Figure 2: *Talk of the Town* logic model and mechanisms of change



Timeline

The timeline, workflow and milestones of the project are presented in Table 2. The evaluation had two significant pinch points. These were during recruitment and subsequent randomisation, and during the post-testing period when schools were undertaking national tests and preparing students for school transition in Year 6. It should be noted that the trial was not registered with an external agency.

Table 2: Timeline, milestones and deliverables of evaluation

Date	Milestone	Responsibility*
May 2013	Meet with delivery team Agree equations for analysis Agree measures	AT, LO, CR, TCT delivery team
June 2013	Recruit schools	TCT delivery team with support from AT
June–Sept 2013	KS1 results used to assign to condition	AT
June 2013	Recruit RA	AT, LO with input from delivery team on specification
Sept–Dec 2013	Pre-intervention testing	RA, AT, LO, Admin
Jan–June 2014	Pre-test data set created	RA, AT
Sept–Oct 2014	Process data collected (interviews undertaken)	RA, AT
April–June 2015	Post-test attainment measures undertaken Questionnaire issued and completed	RA, AT, LO, Admin
June–July 2015	Final data set created	RA, AT, LO, CR
Aug–Sept 2015	Data analysis undertaken	CR, AT
November 2015	Final report completed and submitted to EEF	AT, LO

**AT-Allen Thurston, CR-Cary Roseth, LO-Liam O'Hare, RA-Research Assistant, Admin-Administrator to project*

Costs

The intervention was implemented at a whole-school level in 31 schools. The mean number of students per intervention school was 259.86. Costs calculated as per guidance from the EEF were as detailed below. The total cost of implementing the *Talk of the Town* in 31 intervention schools excluding any evaluation and overheads was £615,867.50. This figure was obtained from the Educational Endowment Foundation Grant Award Ledger. It included staff costs at an estimated 25% of the grant award total (estimated by The Communication Trust as being those staff costs that were due to delivery, but not development of *Talk of the Town* or running of the RCT) and included coordination costs (£88,661.75) staff expenses for SaLTs (£42,502), staff salaries for SaLTs and organisers (£430,184), cost of materials (£44,530) and overheads (£9,989.75). Intervention schools made a contribution to these costs totalling £105,336.

The yearly cost per school of the programme was £13,244.46 (calculated by converting the cost over 18 months to a cost over 12 months by multiplying by 2 and dividing by 3, and dividing that cost by the number of schools which was 31) and a cost per student of £50.97 (obtained by dividing the cost per school by the mean number of pupils per school, which was 259.86). Note that all students in the school accessed support through TotT at a universal level, with some students also receiving targeted support, but only those in Year 3/6 at the end of the trial were tested on the NGRT.

Impact Evaluation

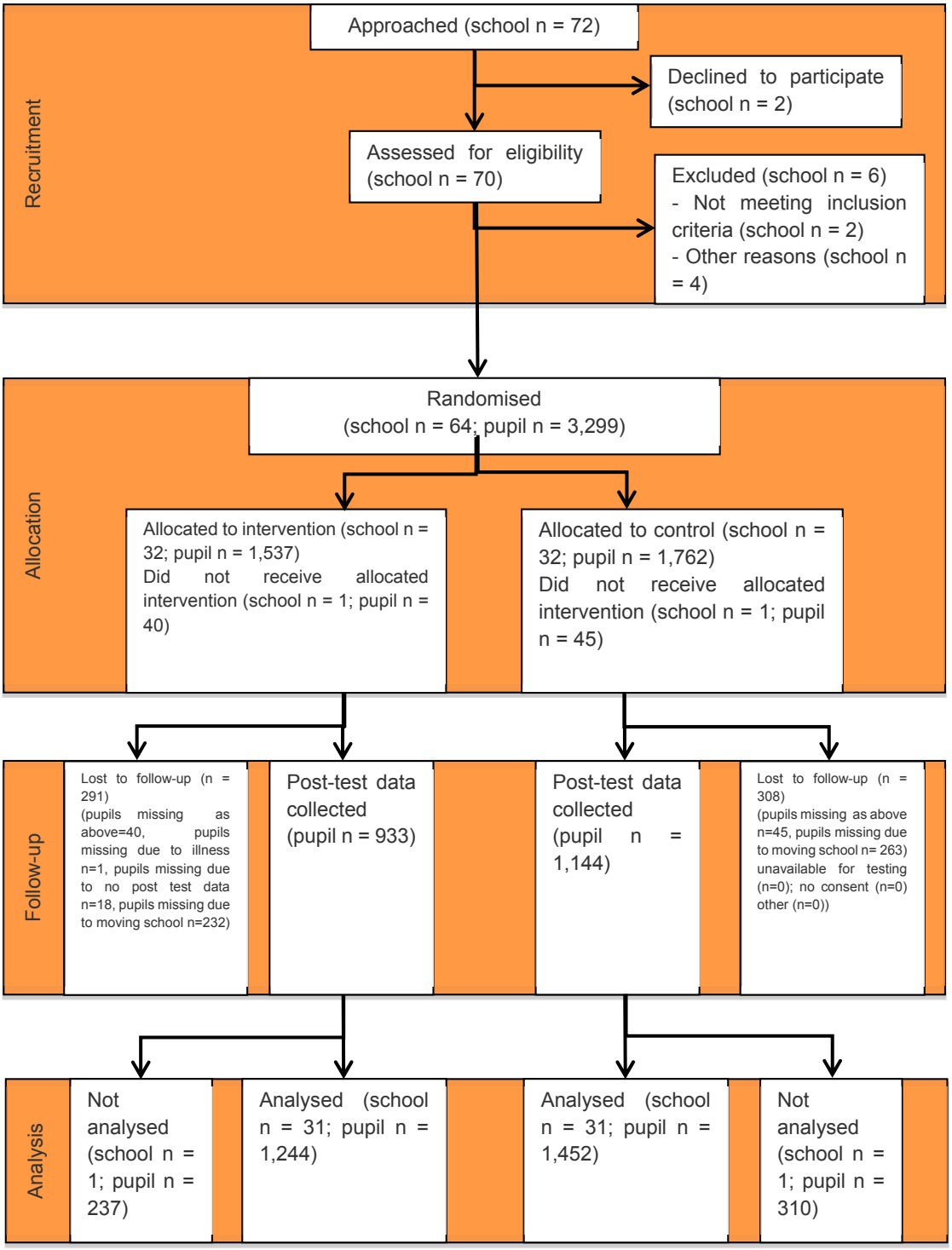
Participants

Figure 3 summarizes participant flow by condition for the primary outcome.

During the allocation phase, one school dropped out after pre-test due to a change of headteacher. During the follow-up phase, one school completed both pre and post-tests, but failed to arrange for collection by Granada Learning. This left a total of 62 schools that completed pre and post-tests that were analysed as intention to treat. The 'pair' from randomisation was not lost to the trial.

Within the original 64 schools there were 3,299 students who completed NGRT pre-test, and 623 of these students were missing at post-test. One student was missing through illness, 40 students were missing as a school failed on three occasions to open the door to the courier booked to pick up the post-test booklets and then closed for the summer holidays, one school in the control arm withdrew and the other 514 students missing at post-test were missing due to the fact that they had left the school. The low missing-by-absence rate was because of low absence on the day selected for testing, and the fact that individuals could undertake the test before it was sent to GL for processing. Data is presented in Figure 3 in respect of participant flow. For the ACE measure, 374 students completed the pre-test and 294 completed the post-test. The reason for attrition in the post-test ACE results is as follows: left school n=52, absent on testing dates n=22, school's inability to provide students for testing n=5, refused consent to test n=3. In the protocol we stated that *'If attrition of schools is low, as we expect, we will use case deletion, but if not, we will use multiple stochastic regression imputation for missing outcome data (see Puma et al., 2009).'*

Figure 3: Participant flow diagram



Minimum detectable *ES*s

Table 3 reports the minimum detectable *ES*s at different stages in the trial.

Table 3: Minimum detectable *ES* at different stages

Stage	N [schools/ pupils] (n= intervention ; n=control)	Correlation between pre-test (+other covariates) & post-test	ICC	Blocking/ stratification or pair matching	Power	Alpha	Minimum detectable *ES* (MDES)
Protocol	62 / 1,550 (31; 31)	0.60	.07	School blocking, 31 schools	80%	.05	0.082
Randomisation	64 / 3,299 (32; 32)	0.60	.07	School blocking, 32 schools	80%	.05	0.084
Analysis (i.e. available pre- and post-test)	62 / 2,696 (31; 31)	0.75	.06	School blocking, 31 schools	80%	.05	0.084

Pupil characteristics

School-level baseline comparisons are presented in Table 4. Overall, the study involved 3,299 students from 64 schools, with an average of 25.98 students per school (SD = 8.23, range = 12 to 94). There were slightly fewer students in *Talk of the Town* condition (n = 1,537) than in control (n = 1,762), a difference in sample size that may be partly attributed to the size of the different schools assigned to each condition. In the control condition, n = 7 (22%) schools enrolled ≥ 60 students each, while only n = 1 (3%) *Talk of the Town* schools enrolled ≥ 60 students. Thus, it was not surprising that the average number of students enrolled in control schools was slightly larger than in the *Talk of the Town* schools, $F(1,125) = 4.73$, $p < .05$, $\eta_p^2 = .03$.

Table 4: School-level baseline comparisons

Variable	Intervention group		Control group	
School level				
Categorical variables				
Number of students per school	**n (missing)**	**Percentage**	**n / N (missing)**	**Percentage**
< 40	5 (0)	15.6	3 (0)	9.37
40–49	12 (0)	37.5	9 (0)	28.13
50–59	14 (0)	43.75	13 (0)	40.63
60–69	1 (0)	3.13	5 (0)	15.6
70–79	0 (0)	…	0 (0)	…
80–89	0 (0)	…	1 (0)	3.13
90–99	0 (0)	…	0 (0)	…
> 100	0 (0)	…	1 (0)	3.13
Totals	32 (0)	100	32 (0)	100
Continuous variables	n (missing)	Mean	n (missing)	Mean
Sample size	1535 (2)	24.53	1760 (2)	27.37

Student-level baseline comparisons are presented in Table 5. Overall, demographic characteristics were evenly distributed across the *Talk of the Town* intervention schools and control conditions, with no categorical variable differing by more than 2% and no evidence of any statistically significant differences. However, there was evidence of a statistically significant difference between the *Talk of the Town* intervention condition and control conditions in baseline (i.e, pre-test) NGRT scores, the primary outcome measure, $F(1,3293) = 4.13$, $p = .04$, $\eta_p^2 < .001$. But the magnitude of this mean difference was very small, with less than 1% of the variance in NGRT pre-test scores accounted for by intervention condition, and there was little difference in the amount of variability within each condition (e.g. standard deviations). Thus, sample size rather than a meaningful difference in pre-test scores accounted for this statistically significant mean difference. It should be noted that any pre-test differences were taken into account in analyses.

Table 5: Student-level baseline comparisons

Variable Student level Categorical variables	Intervention group n/N (missing)	 Percentage	Control group n/N (missing)	 Percentage
Female	741 / 1537 (0)	48.2	844 / 1762 (0)	47.9
Male	796 / 1537 (0)	51.8	918 / 1762 (0)	52.1
FSM eligible	961 / 1537 (2)	62.6	1132 / 1761 (1)	64.3
Not FSM eligible	574 / 1537 (2)	37.4	629 / 1761 (1)	35.7
Not SEN	1156 / 1536 (1)	75.3	1350 / 1717 (45)	78.6
SEN	380 / 1536 (1)	24.7	367 / 1717 (45)	21.4
EAL	1394 / 1534 (3)	90.9	1573 / 1762 (0)	89.3
Not EAL	140 / 1534 (3)	9.1	189 /1762 (0)	10.7
Minority	1307 / 1534 (3)	85.2	1461 / 1753 (9)	83.3
Not Minority	227 / 1534 (3)	14.8	292 /1753 (9)	16.7
Continuous variables	**n (missing)**	**Mean (SD)**	**n (missing)**	**Mean (SD)**
NGRT score - Pre-test	1535 (2)	95.87 (12.56)	1760 (2)	96.76 (12.36)

Outcomes and analysis

Primary outcome: NGRT reading scores

Primary analysis. The first set of analyses compared the relative effect of *Talk of the Town* and control conditions on the primary outcome, NGRT reading scores. Results of the unconditional LMM indicated that NGRT scores varied significantly between schools, τ_{00} = 11.67, 95% CI = 6.03 to 17.31, and among students within a school, σ^2 = 180.77, CI = 171.01 to 190.53. This corresponds to an intraclass correlation (ICC) of 0.06 and design effect of 3.49. Thus, even though only 6% of the NGRT variance occurred across schools, there was compelling evidence of the need to account for clustering within schools using multilevel models (cf. Muthén & Satorra, 1989, 1995).

Results of the conditional LMM (i.e. ANCOVA) model indicated there was no significant difference between *Talk of the Town* and control schools after controlling for pre-test reading scores, γ_{10} = -0.46, CI = -2.42 to 1.50, Hedges's *g* = -0.03, CI = -0.46 to 0.40 (*Talk of the Town*:

n = 38.88/school, M = 99.49, SD = 14.79; Control: n = 45.38/school, M = 99.97, SD = 13.02). Thus, there was *no* evidence that the *Talk of the Town* intervention influenced mean post-test NGRT scores. There was, however, clear evidence that residual variability among students was greater in *Talk of the Town* schools, σ^2 = 118.15, CI = 108.74 to 127.56, than in control schools, σ^2 = 77.15, CI = 71.49 to 82.81, after controlling for pre-test NGRT scores. In fact, there was 53.1% more variability in *Talk of the Town* schools compared to control schools. Table 6 summarizes the mean comparison between conditions and Table A1 in the Appendix B reports the model parameters and fit statistics.

Table 6: Primary analysis of primary outcome, the NGRT readings scores

	Raw means				ES		
	Intervention group		Control group				
Outcome	n (missing)	Mean (95% CI)*	n (missing)	Mean (95% CI)*	n in model (intervention; control)	Hedges g (95% CI)	p-value
NGRT post-test	1244 (293)	99.49 (98.67, 100.31)	1452 (310)	99.97 (99.30, 100.64)	2696 (1244; 1452)	-0.03 (-0.46, 0.40)	n.s.

Sensitivity analysis. Mean comparisons revealed that, on average, students missing NGRT post-test scores had lower NGRT pre-test scores, $F(1, 3297) = 8.28$, $p < .01$, $\eta_p^2 < .01$, and that Year 2 students were more likely to have missing NGRT post-test scores than Year 5 students, $\chi^2(1) = 22.05$, $p < .001$. Accordingly, we tested whether the parameter estimates were sensitive to missing data by estimating the LMMs again using multiple imputation. The basic premise is that similarity between the results based on the observed and multiply imputed data strengthens confidence that parameter estimates were not sensitive to missing data.

As reported in Table A1, there was little difference between the results based on the obtained and multiple imputation data. This strengthens confidence in the statistical validity of the results and the assumption of an ignorable missing data mechanism (e.g. MAR) for the NGRT analyses.

Student covariates. Next, we examined the association between student-level predictors and NGRT post-test scores (see Table A1 for model parameter estimates). Results again indicated that there was no significant difference between *Talk of the Town* intervention schools and control schools, γ_{10} = -0.43, CI = -2.33 to 1.47, even after controlling variance accounted for student covariates (i.e. gender, year, minority status, eligibility for FSM, and EAL and SEN status). Controlling for student covariates reduced between-school variance by 5.6%, and within-school variance by 3.8% in *Talk of the Town* schools and 2.4% in control schools. Post-test scores were significantly lower among students who received free school meals, γ_{05} = -1.03, CI = -1.81 to -0.25, Hedges's *g* = -0.08, CI = -0.52 to 0.37 (FSM: n = 30.66/school, M = 96.90, SD = 13.45; No FSM: n = 53.59/school, M = 101.38, SD = 13.84). Post-test scores were also significantly lower among students receiving special education (SEN), γ_{07} = -4.19, CI = -5.13 to -3.25, Hedges's *g* = -0.33, CI = -0.83 to 0.18 (SEN: n = 19.94/school, M = 90.44, SD = 12.63; No SEN: n = 64.31/school, M = 102.64, SD = 12.93). There was no evidence that mean post-test scores were related to students' gender, year in school, or minority or EAL status.

FSM subgroup analysis. In keeping with EEF's policy on analysis (2015), we computed an *ES* on the primary outcome for FSM pupils using a separate model. Table 7 summarizes the mean comparison between conditions and Table A2 reports the model parameters and fit statistics.

Consistent with the overall sample, the results of the unconditional LMM analysis indicated that the NGRT scores varied significantly across students receiving FSM within a school, σ^2 = 166.45, CI = 151.30 to 181.60, and between schools, τ_{00} = 13.51, CI = 5.12 to 21.90. This corresponds to an intraclass correlation (ICC) of 0.08 and a design effect of 2.34.

Table 7: FSM subgroup analysis of NGRT readings scores

	Raw means				ES		
	Intervention group		Control group				
Outcome	n (missing)	Mean (95% CI)*	n (missing)	Mean (95% CI)*	n in model (intervention; control)	Hedges g (95% CI)	p-value
NGRT post-test	574 (2)	96.36 (95.16, 97.56)	629 (1)	97.35 (96.39, 98.31)	1203 (574, 629)	-0.09 (-0.73, 0.55)	n.s.

Also consistent with the overall sample, there was no significant difference in the mean NGRT scores of students receiving FSM in *Talk of the Town* and control schools, γ_{10} = -1.17, CI = -3.46 to 1.12, which corresponds to an *ES* of Hedges's *g* = -0.09, CI = -0.73 to 0.55 (*Talk of the Town*: n = 17.94/school, M = 96.36, SD = 14.68; Control: n = 17.94/school, M = 97.35, SD = 12.30). There was, however, clear evidence that residual variability among students was greater in *Talk of the Town* schools, σ^2 = 125.48, CI = 108.60 to 142.36, than in control schools, σ^2 = 78.55, CI = 68.83 to 88.27, after controlling for pre-test NGRT scores. In fact, there was a 59.7% more variance in NGRT scores in *Talk of the Town* schools than in control schools.

NGRT predicted by language (ACE) subgroup. As specified in the protocol, we also examined the predictive association between language (as measured by ACE) and post-test NGRT scores. We emphasize that these analyses involve only the subgroup of Year 2 students who completed the ACE measure. Table 8 summarizes the mean comparison between conditions and Table A3 reports the model parameters and fit statistics.

Consistent with the overall sample, the results of the unconditional LMM indicated that NGRT scores varied significantly across students, σ^2 = 90.19, CI = 73.92 to 106.46, and between schools, τ_{00} = 16.82, CI = 3.67 to 29.97. This corresponds to an ICC of 0.16 and a design effect of 1.58. Compared to the overall sample, this suggests that there was approximately twice as much variation across schools in NGRT scores among the subsample of students who completed the ACE measure. However, this finding should be interpreted with caution as the number of students per school in this subsample was quite small (n ~ 4.5 to 4.88 per school) and the design effect, or multiplier needed to be applied to the stand errors to correct independence violations in nested data, was only 1.58. In fact, according to some researchers (e.g., Muthén & Satorra, 1989, 1995; Peugh, 2010), design effect estimates less than 2.0 indicate that multilevel modelling is not needed.

Table 8: Supplementary analyses of NGRT readings scores predicted by ACE language scores

Outcome	Raw means				ES		
	Intervention group		Control group				
	n (missing)	Mean (95% CI)*	n (missing)	Mean (95% CI)*	n in model (intervention; control)	Hedges g (95% CI)	p-value
NGRT post-test	144 (47)	92.28 (90.22, 94.34)	156 (30)	93.07 (91.52, 94.62)	300 (144; 156)	-0.06 (-1.34, 1.22)	n.s.

Also consistent with the overall sample, the results of the conditional LMM indicated there was no significant difference between *Talk of the Town* and control schools after controlling for pre-test reading and language scores, γ_{10} = -0.70, CI = -3.70 to 2.30, Hedges's *g* = -0.06, CI = -1.34 to 1.22 (*Talk of the Town*: n = 4.50/school, M = 92.28, SD = 12.64; Control: n = 4.88/school, M = 93.07, SD = 9.90). There was, however, clear evidence that residual variability among students was greater in *Talk of the Town* schools, σ^2 = 118.37, CI = 88.30 to 148.44, than in control schools, σ^2 = 64.61, CI = 48.60 to 80.62, after controlling for pre-test reading and language scores. In fact, there was 83.2% more variability in *Talk of the Town* schools compared to control schools.

In terms of missing data, mean comparisons revealed that, on average, students with missing ACE post-test scores had lower ACE pre-test scores, $F(1, 370) = 4.78$, $p = .02$, $\eta_p^2 = .01$. Nonetheless, there was little difference between the fixed-effect parameter estimates based on the obtained data and the multiply imputed data (Table A3). Again, this strengthens confidence in the statistical validity of the results and the assumption of an ignorable missing data mechanism.

Secondary outcome – ACE language scores

Primary analysis. The next set of analyses compared the relative effect of *Talk of the Town* and control conditions on the secondary outcome, ACE language scores. Results of the unconditional LMM indicated that ACE scores varied significantly across students within a school, σ^2 = 91.49, CI = 81.73 to 101.25, and between schools, τ_{00} = 34.24, CI = 14.70 to 53.78. While this corresponds to an ICC of 0.27, the relatively small average number of students per school who completed the ACE post-test (n_c = 292 / 64 = 4.56) resulted in a relatively small design effect of 1.97.

As summarized in Table 9 (see Table A4 for complete model), results of the conditional LMM (i.e. ANCOVA) model indicated that there was no significant difference between the *Talk of the Town* and control schools in mean post-test ACE scores, γ_{10} = 1.42, CI = -2.21 to 5.05, Hedges's *g* = 0.11, CI = -1.19 to 1.41 (*Talk of the Town*: n = 4.41/school, M = 92.41, SD = 14.28; Control: n = 4.72/school, M = 91.42, SD = 12.49). There was, however, clear evidence that residual variability among students was greater in *Talk of the Town* schools, σ^2 = 106.57, CI = 78.27 to 134.87, than in control schools, σ^2 = 78.13, CI = 58.63 to 97.63, after controlling for pre-test ACE scores. In fact, there was 36.4% more variability in ACE post-test scores in *Talk of the Town* schools than in control schools. Thus, while there was little variability between *Talk of the Town* schools, students within these schools varied significantly in the effect of the intervention on post-test ACE scores, especially compared to students in control schools.

Table 9: Primary analysis of the secondary outcome, ACE language scores

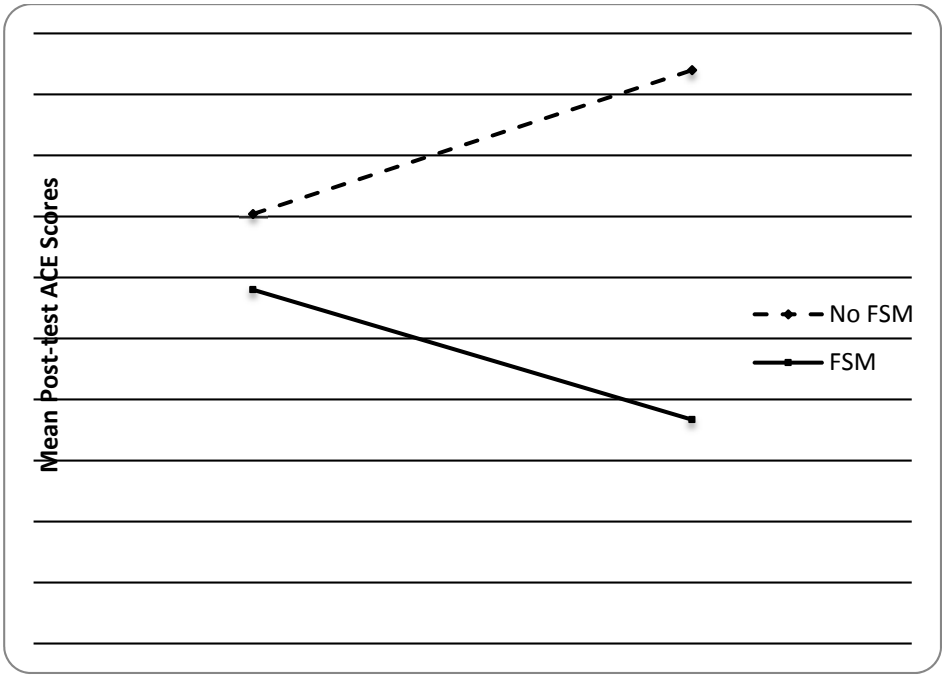
Raw means	ES

Outcome	Intervention group		Control group				
	n (missing)	Mean (95% CI)*	n (missing)	Mean (95% CI)*	n in model (intervention; control)	Hedges g (95% CI)	p-value
NGRT post-test	141 (47)	92.41 (90.05, 94.77)	151 (35)	91.42 (89.43, 93.41)	292 (141; 151)	0.11 (-1.19, 1.41)	n.s.

Sensitivity analysis. Again, there was little difference between the fixed-effect parameter estimates for the post-test ACE scores based on the obtained and multiple imputation data, strengthening confidence in the results (see Table A4).

Student covariates. Interestingly, the results of the student covariate model indicated that the difference in mean post-test scores between the *Talk of the Town* intervention schools and control schools depended on FSM status, γ_{20} = -6.59, CI = -11.55 to -1.63 (see Table A4). As displayed in Figure 4, this suggests that the *Talk of the Town* intervention had a positive effect on post-test ACE scores compared to control schools among students who did *not* receive FSM, but a negative effect on the scores of students who did receive FSM in the small sub-sample used in this study. For *Talk of the Town* schools, this effect corresponded to a Hedges's *g* = 0.42, CI = -1.56 to 2.40 (No FSM: n = 2.88/school, M = 94.40, SD = 13.38; FSM: *n* = 1.53/school, *M* = 88.67, SD = 15.28).

Figure 4: Observed mean post-test ACE scores by condition and students' FSM status



Results also indicated that mean post-test ACE scores increased with students' reading age, γ_{04} = 0.46, CI [= 0.21 to 0.72], Hedges's *g* = 0.35, CI -0.95 to 1.64 (*Talk of the Town*: n = 4.41/school; Control: n = 4.72/school; Reading age SD = 5.01). There was no significant effect on students with SEN γ_{07} = -2.45, CI = -5.17 to 0.27, but this finding should be interpreted with caution because it was not replicated in the multiple imputation analysis.

Adding the student-level covariates reduced between-school variance in ACE language scores by 21.8%, and within-school variance by 4.1% in *Talk of the Town* schools and 3.9% in control schools. There was no evidence that ACE scores were related to students' gender, minority, or English as an additional language (EAL) status.

FSM subgroup analysis. In keeping with EEF's policy on analysis (2015), we used a separate model to compute an *ES* on the secondary outcome for FSM pupils. Table 10 summarizes the mean comparison between conditions and Table A5 reports the model parameters and fit statistics.

As with the overall sample, there was no evidence that ACE post-test scores among the FSM subgroup varied significantly between schools, τ_{00} = 23.40, CI = -6.59 to 53.39. But there was clear evidence of variability among students receiving FSM, σ^2 = 106.99, CI = 73.00 to 140.98. While this corresponds with an ICC of 0.18, the relatively small average number of students receiving FSM with ACE post-test scores (n_c = 125/64 = 1.95) only resulted in a design effect of 1.17. Once again, given the small number of students per school in this subsample and the small design effect, the amount of variability attributed to between-school differences should be interpreted with caution.

Table 10: FSM subgroup analysis of ACE language scores

Outcome	Raw means				ES		
	Intervention group		Control group				
	n (missing)	Mean (95% CI)*	n (missing)	Mean (95% CI)*	n in model (intervention; control)	Hedges g (95% CI)	p-value
ACE post-test	49	88.67 (84.28, 93.06)	76	90.80 (87.88, 93.72)	125 (49, 76)	-0.23 (-2.27, 1.81)	n.s.

Consistent with the overall sample, there was no significant difference in mean ACE post-test scores of students receiving FSM in *Talk of the Town* and control schools, γ_{10} = -3.09, CI = -7.87 to 1.69, Hedges's *g* = -0.23, CI = -2.27 to 1.81 (*Talk of the Town*: n = 1.53/school, M = 88.67, SD = 15.28; Control: n = 2.38/school, M = 90.80, SD = 12.77). There was, however, clear evidence that residual variability among students was greater in *Talk of the Town* schools, σ^2 = 141.07, CI = 76.39 to 205.75, than in control schools, σ^2 = 87.46, CI = 53.45 to 121.47. In fact, there was 61.3% more variance in ACE scores in *Talk of the Town* schools compared to control schools. Here again, this suggests that while there was little variability between *Talk of the Town* schools, students within these schools varied significantly in the effect of the intervention on post-test ACE scores, especially compared to students in control schools.

Process evaluation

As previously stated the process evaluation aimed to assess whether conditions necessary for successful implementation of TotT were achieved, namely:

- whether a whole-school approach was achieved;
- whether targeted interventions were delivered;
- whether targeting of interventions at specific students was appropriate;
- what barriers existed to achieving the required conditions for optimal implementation; and
- the opinions of teachers, members of senior management teams and local authority representatives/managers regarding how worthwhile/attractive TotT was as an intervention.

Was a whole-school approach achieved?

Data from the survey indicated that the majority of respondents reported that they were happy that their school had adopted the programme (90% agreed/strongly agreed) and a similar number were 'enthusiastic' about the approach (>86% agreed/strongly agreed). Fewer than 5% reported that the working environment was not conducive for the approach taken by the programme, and more than four in five teachers felt they had fully engaged with the programme.

Were targeted interventions delivered?

Resources were received well and shared among teachers and classes. There was evidence that this led to use of resources throughout the school. The programme was seen as efficient, and a number of teachers reported they were pleased with and appreciative of the TotT intervention. Teachers reported feeling supported and "listened to", and that the sharing of techniques between schools was beneficial. In particular, they noted the following:

Resources—Four SMTs and four teachers reported that the *Talk of the Town* resources were excellent. Resources were seen as containing information to run the groups well, and any problems with running the group sessions could be solved when the SaLT attended each week.

Word of the week (WOW)—This was reported as being of great benefit to children's vocabulary, and was positively received by children. WOW was seen as a useful technique, which some teachers planned to expand upon. One SMT commented that the school was using the technique across all classes, and there were plans to integrate it into assemblies.

Collaboration with SaLT—A positive relationship was reported by all SMT and teachers interviewed. SaLTs seemed to be well integrated into the school, with multiple mentions of them being "approachable" and a close working relationship across the board.

Was targeting of targeted intervention at students appropriate?

The intervention was reported to be benefitting communication, confidence and behaviour. Teachers "hoped" that the reading and writing skills would improve in the long term, but stopped short of any claims that they had already done so.

Vocabulary—There was a widespread feeling that the strategies were of benefit to both children's vocabulary, and to teachers' identification of children with poorer vocabulary and subsequent targeting of the development of more fluent speech in these children.

Communication—Another widely reported benefit was that communication, 'story talk' and expression had been improved.

Literacy—Teachers reported that language benefits should improve literacy, e.g. ability to decode new words. Numerous schools reported that they "hoped" to see improvement in reading and writing, i.e. that these would improve long term. Improvements were reported in vocabulary and speech that seemed to be instantly noticeable.

Confidence—Confidence boosts were reported in numerous schools, e.g. previously quiet children engaged more in speech, increased their participation in conversation during recreation breaks, and pupils showed reduced fear of getting an answer wrong.

Behaviour—Many SMTs and teachers reported that improved listening skills had improved classroom behaviour.

However, data from the survey indicated that one in ten teachers felt that the materials could have been designed to be more suitable for individual student needs. This was a fact highlighted by the interview data.

It was reported that schools had implemented Talk Boost in a manner at odds with TotT's planned strategy. Although TCT have indicated that the same person should deliver Talk Boost, some schools got the impression from training that different teachers can deliver the targeted intervention. When this happened they reported that this led to repetition and difficulty in ensuring efficient delivery. This does question whether a whole-school approach was achieved and there may need to be some thought given to how to optimise whole-school planning. This indicated that there were some instances where training had not resulted in optimal implementation of the targeted intervention.

> *"It was suggested that [Talk Boost] … didn't have to be the same person every time that delivered it, but actually when we done that, when we did it we found it really difficult…. With 3 of us delivering it we found we were doing the same activities quite a lot as it repeats"* Teacher 1

Some teachers reported that the four interventions were not at the optimum level to promote learning. Teachers felt that resources were not age-appropriate. Teachers from both KS1 and KS2 felt that the resources were too basic and did not challenge some of the more able pupils:

> *"The phonics bingo … I do find we are having to adapt those as they are very, very basic for my children"* Teacher 3

> *"It's the follow-on activities, I don't think they are working so well because other Y2's and some of the things they are given to do for my better Y2 they do not interest them, they are not stretching them"* Teacher 3

Some teachers reported that materials were overly focused on KS2 with little to support lower ability learners and KS1 pupils. This led to materials being too challenging and not appropriately differentiated for low-ability students. Teachers indicated that matching differentiated materials

to students was a matter of matching the academic level and matching the age-appropriate interest level of the materials:

> *"Some of your resources are very KS2 focused and don't take into account the lower end, especially with some of the resources to do with vocab our SaLT actually sent us a more friendly KS1 resource that she uses"* Teacher 2

> *"…for example with the talking hats, they have the potential to be used in KS1 but again needs adaption to be used down there, so that is just a practical application of the resources"* School Manager 1

> *"One of the group work activities is more based towards KS2, and a lot of them I feel are actually a bit old for our children, particularly the group work where they all have roles. I would like something designed for KS1"* Teacher 4

Fidelity: What barriers existed to optimal implementation?

There was an overall feeling that the project led to pressures on time, staffing and room use. Effective implementation of strategies in the classroom and a shift in attitude about the importance of speech for reading and writing are seen as needing more time in a crowded curriculum.

Time—One of the most widely reported issues with implementation was time pressure. Finding the time to implement the intervention was a frequently reported problem, and there was an overall feeling of it being rushed. The timing of introduction of some parts of the targeted intervention at the beginning of the summer term was problematic. The summer term was described as a "mani'" term by more than one teacher and may be a difficult term to engage teachers effectively. This was probably due to SATs and preparations for school transfer at the end of Year 6. A number of teachers reported that delivering a complex intervention with four targeted interventions led to pressures on time and may have led to some implementation integrity issues:

> *"The difficulties …is staffing of everything and the finding of rooms, somewhere which is nice and quiet"* School Manager 2

> *"If we had more time and more staff* (we would) *be able to put more children onto the Talk Boost programme"* Teacher 5

> *"Everyone knows TotT is a priority, … and therefore it must happen, but if you've got staff who are ill, you've got children who are not there, something comes up and a teaching assistant needs to cover the teacher or take the class … your intervention can easily go by the wayside … – that is the biggest issue"* School Manager 3

> *"Timetabling and just pressures of time, we have so many subjects to teach and we have got our own pressures from Ofsted … it has been a real struggle"* School Manager 4

A significant number of respondents agreed/strongly agreed that other activities got in the way of delivering the programme from the survey (58% agreed/strongly agreed) indicating that there was risk of the programme being displaced by other school activities. This emphasises the time constraints most teachers felt they were working under.

Resources inherent to the school—Another issue reported with frequency was the problem with finding staff and rooms available for TotT interventions. Teachers reported that it was difficult to ensure staff and rooms were available at the same time, and these limitations were seen as a problem for many schools. One reported that constructing a timetable alleviated the room issue.

> *"Staffing – it's just finding enough staff because … we don't always have the staff we would like to have to run all these interventions"* School Manager 5

Heavy workload and pace of implementation—The workload was reported as being intense or heavy by five SMTs and they also reported that it was not ideal for the SENCO to be the Communication Lead. Making so many changes in a short period of time was difficult for schools, and it may take a long time for changes in school policy, timetabling, teaching and learning to be efficiently embedded. Teachers reported that being provided with such a large amount of information, and with so many different strategies in rapid succession, was problematic. More gradual introduction to the different strategies would allow teachers to become more familiar with them, and implement them more effectively at the start of the project.

> *"My number one thing I have asked for is let's just slow down a little and embed what we are doing"* Teacher 6

> *"The scale at which you introduce the elements would benefit from being extended, in terms of instead of having the second year to be able to consolidate, maybe look at doing two elements one year and two elements the second year because I think that is the thing that is coming up, time, being able to give enough time to address it as it has been strategy after strategy to the extent you think what have I been told as it has come so quickly"* Teacher 7

> *"I feel like (we) have been thrown a lot of information this year … every six weeks something new gets given to us to do, we haven't had the chance to fully embed it, we have that many other things going on"* Teacher 8

Early identification and progression tool use (at beginning)— It was reported that it was difficult to get used to the idea that speech is important for reading and writing, and the attitude shift is taking time.

Formative findings: Suggested improvements to Talk of the Town

More children to take part in intervention—One teacher reported that a larger group of children taking part in the targeted interventions would have been a benefit, as teachers were aware of many more children who would have benefitted from it.

> *"We would like to have sent more children in the intervention group, em I think, I know the reasons why it is down to 4, but within our classrooms like, a nice group of 6 would have been nice to send out 'cause we felt like we had so many children we wanted to send on it"* Teacher 9

Better organisation of the interventions—Earlier planning of the intervention would have been appreciated. The onset of the project was seen as too sudden, and teachers were unsure of what they were doing; it felt rushed and "a rollercoaster" at the beginning. The late beginning of the KS2 aspect of the project was seen as a problem.

> *"Perhaps the only thing is if it could be plotted when interventions are happening a little bit earlier on so that we know when everything is taking place"* Teacher 10

> *"Bit of a rollercoaster ride at the beginning, when we came to the project it was as if all of a sudden let's go and we were floundering a bit and we really weren't sure … of what we are doing. …I do feel the first term was very rushed"* School Manager 6

> "*We did struggle with the KS2* (targeted intervention) *coming out so late. I know we haven't been able to get through the whole programme with those children … having that information earlier would have been better*" School Manager 7

Adaptations to resources—As noted above, teachers reported that they would like packages to be specifically designed with 'age interest level' built into the materials for both KS1 and KS2. This would optimise the ability to use KS1 work for students in Year 3 and above and KS2 work in Years 1 and 2 if appropriate. More detailed observation sheets would be appreciated, i.e. for Talk Boost with more examples of what to look out for.

> *"In the Talk Boost intervention, more focused observation sheets for the children, so more examples of what we could look out for each childn on the five different areas they are working on"* Teacher 12

A number of staff suggested that the posters could be made more child-friendly, as they were quite 'bland'. The group work activities were seen as being pitched at older children, and biased towards KS2, so weren't as suitable for KS1 as they could have been.

Opinions of teachers, members of senior management teams and local authority representatives/managers

Survey results indicated that most respondents (94.2%) agreed/strongly agreed that the programme engaged and motivated the children and 80% said they felt engaged and excited with the programme themselves. Nearly every respondent (98%) felt that senior management were keen to see the programme work, but this is not surprising at the 'opt-in' to the project was made at SMT level.

Many of the survey respondents felt the programme was an important addition to the provision in their school (88%) and that they would be happy for the school to continue with TotT in subsequent years (82%).

Control group activity

As was noted earlier, and stated in the published protocol, there was no systematic plan to collect data from control group schools as part of this evaluation.

Conclusion

Limitations

There are a number of limitations to the study. The protocol clearly stated that there was a plan to develop the model for analysis after initial planning. Analysis undertaken offered best statistical processes and was planned to take account of sample composition properties. However, the analysis model was not in place before the study was undertaken. The counter-measure to this is that an academic who was out of the country and had no contact with the delivery team developed the model. This academic also had no access to condition of the sample and so model building and analysis were undertaken blind to condition. Risk of bias is therefore unlikely.

The level of individual attrition due to movement of students with low literacy out of schools was not anticipated and is an obvious limitation of the evaluation. Strengthening confidence in the results, however, multiple-imputation and maximum likelihood produced very similar parameter estimates and standard errors. This suggests that the results were not sensitive to any systematic bias that might be related to attrition among students with low pre-test literacy scores, or any of other missing data mechanism related to the other auxiliary variables included in the imputation process (e.g. gender, year, minority status, FSM, EAL, SEN). Thus, even as it ultimately remains unverifiable as to whether another unmeasured cause of missingness may be related to differential outcomes, we consider these findings to be statistically robust.

Moving beyond the question of whether attrition may have influenced the evaluation's results, it is still an important finding that attrition was greater among students with low pre-test literacy scores. School transfer has long been reported to result in significant decrease in academic motivation and student achievement (e.g. Finger & Silverman, 1966; Petersen & Crockett, 1985; Barber & Olsen, 2004). What appears to have happened in this sample is that a vulnerable group of students, in terms of their literacy performance, is showing increased mobility between schools, which has been shown to result in decreased academic achievement. This is a finding that might be worthy of further investigation. Attrition rates due to move of school were similar to those previously observed in a two-year randomised trial in Fife, Scotland, although this trial had more attrition due to loss of schools from assigned condition (Tymms et al., 2011).

Interpretation

The main finding of this study was that there was no evidence of enhanced reading attainment as measured by the NGRT for the *Talk of the Town* group compared to the control group on the main outcome of reading. The causal mechanisms/logic model for *Talk of the Town* (presented in Figure 2) indicated a number of areas in the anticipated logic model that could explain the failure to observe effects.

The overall logic model of the *Talk of the Town* was to ensure enhanced targeted provision of speech and language interventions, ensuring the correct interventions get to the right students. Some teachers reported that the activities did not challenge the children enough, some that they were too hard. Dockrell & King (2007) have previously reported the importance of matching the level of the intervention to the level of the students to maximise success. It would appear that the matching was not optimal in this instance and therefore thought should be given as to how to ensure the right students are identified and matched to the correct interventions. One of the goals of TotT was to identify language needs. It would appear to be failing in this goal that may lie at the heart of the failure to detect observable effects of the programme. Teachers

reported varying degrees of satisfaction with the targeted intervention of TotT. While the actual targeted interventions were well received and thought to be valuable, there remain some unresolved issues about which students were most appropriately matched to the levels of resource. Allied to this were views expressed by teachers that there was repetition and overlap. When resources were found to lack breadth at KS2, then it was noted that materials from KS1 were used to supplement them. Conversely, it was reported that there was a bias in the materials towards KS2. This would indicate that not all activities were in the zone of learning that would have led to enhanced performance for all children. On the other hand, it may be that the programme failed in its stated aim that teachers would be able to identify the correct children for targeted intervention. This may also account for the fact that materials did not appear to be matched at the correct level to children. This would be a differing frame than that presented by the teachers, and would still be a failing in the programme, but must be considered as a possible explanation. Finer grained analysis of outcome measures may help TCT to determine this.

A frequent comment was that time pressures led to the interventions being rushed and not always implemented optimally. This was reported particularly at the start of the programme. The requirement for a whole-school approach in the logic model may have been compromised during the early days of the implementation owing to the sheer volume of new work/methods of working and targeted intervention that were implemented. Teachers in the process evaluation noted the early stages of implementation as problematic. School reform is difficult to achieve and it may be that an intervention with multiple targeted interventions might need a carefully graduated introduction to ensure that resources (staff, staff time, rooms etc.) are available to allow for optimal implementation.

Modelling did indicate that previous reading scores were an accurate predictor of future reading. This is not surprising as the pre and post-reading test (despite being differing versions of the same test) measured the same constructs of reading (sentence completion and passage comprehension).

There are other explanations for failure to detect a main effect. The lack of significant finding regarding the SAS2 post-test reading scores may be due to not enough time being given between pre and post-test. Pre-tests were completed in October 2013 and post-tests in June 2015. The intervention began in January 2014 and there was evidence that delivery may not have been optimal in the early stages. It may be necessary to track the younger students involved in the intervention through to KS2 test scores to ascertain whether there has been a long-term effect of the intervention.

The findings of this study on the secondary outcome were that there were no significant benefits to the *Talk of the Town* group overall, but significant differences favouring *Talk of the Town* group over control group in ACE outcomes for students not in receipt of FSM. However, there were negative effects observed on ACE outcomes for those in receipt of FSM. The interaction of literacy development with FSM is an interesting finding. It may be that the programme is engaging to those with low literacy from a more affluent background, but not for low literacy students from a poorer socio-economic background. However, it does also illustrate that the link between spoken language and literacy is complex and the causal and predictive mechanisms at play between these two aspects of language may require further study to unpack and analyse effectively. In particular, caution must be attached to these findings given the issues with sample size that are constantly referred to in this report.

It could be speculated that the ACE outcome measure overlaps with the *Talk of the Town* theory or model of outcome change better than the NGRT reading test, which was focused on sentence completion and passage comprehension. However, ACE was sat by a very small number of FSM pupils so any results produced from this subgroup are inconclusive at best.

Further work would be required to explore engagement issues or the way that social groups form and mould academic performance.

Overall it could be concluded that the selection of staff interviewed had some positive views on the *Talk of the Town* when it was up and running. They reported benefits to the learning of children and also benefits to the professional practice of those working in the school. Issues were mainly associated with ensuring that targeted intervention matched the required levels of student attainment. In addition, school reform in the early stages of the project was reported to be an issue. Issues with school reform are important for addressing whole-school change in professional practice in any context. Ways to achieve systematic school reform have been well researched. A study involving the implementation of 395 high schools in the USA found that it was often difficult to change pupil instruction during school reform. One of the major issues in determining the ability of school reform to change student instruction was whether or not the school had systemically subscribed to the proposed reform model (Ravitz, 2010). In general, a headteacher made the decision to participate in *Talk of the Town*. This may have meant that teachers were not aware or had not 'bought into' the project in the early phases. In addition, it has been reported that having too many new initiatives going on at the same time can blur the focus of schools during reform and stepwise introduction of school reform initiatives can be desirable (Bronson, 2010). There was some evidence that teachers felt that the pace of change in the early stages of implementation of *Talk of the Town* was faster than was optimal for them. This may have led to a lag period between implementation of *Talk of the Town* and the changes in professional action of teachers that were eventually reported by teachers. During this initial phase, reform may have been less than optimal and some form of 'collective' buy-in might have prevented this. Furthermore, a more staggered or staged approach to implementation might have enabled embedding of programme components before progressing to the next component. This might have alleviated genuine time constraints on teachers in terms of workload.

Another key issue, even when buy-in was achieved, was finding the time to implement TotT. The intervention included four targeted elements, as well as ongoing universal provision and support. Many of the targeted interventions involved three, 30-minute sessions per week. Teachers noted how difficult it was to be able to ensure that there was full implementation of a resource intensive intervention and this may have had a significant effect on implementation integrity.

The size, geographical spread and duration of the study probably give a good indication of how *Talk of the Town* could generalise to schools in England. However, this must be tempered with issues of self-selection of schools, limitation to three distinct geographical locations and the fact that schools had a higher than mean reported FSM percentage. After some initial issues getting the project up and running the teachers reported good implementation. The size and scale of the implementation and the mutualisation of the intervention would lead one to conclude that it had potential to be implemented according to specification. The duration between pre and post-test gave well over 17 months for effects of the project to develop; the project was therefore mature and embedded before post-testing took place. The 64 school design (notwithstanding the 2 school drop-out, leaving 62 schools at post-test) achieved the planned sample. Assessing the implementation across 64 schools in three very different geographical locations across England allows one to conclude that the results are representative and not unduly influenced by local/school differences (security on this statement can also be provided from the HLM analysis looking for school or 'clustering' effects). Loss of two schools was not desirable, but in an RCT of this size was not unexpected. Some attrition at the school level was anticipated, hence the 'over-recruitment' to allow for some leeway should this happen.

In conclusion, it should be noted that there was no evidence of *Talk of the Town* raising reading levels as measured by NGRT on the timescale measured by this study. The explanation of this may be that 1) the *Talk of the Town* programme does not raise levels in reading in terms of sentence completion and passage comprehension; and 2) the *Talk of the Town* programme requires more time to raise levels in reading in terms of sentence completion and passage comprehension. The lack of significant findings could be due to mixed success in implementation during the course of the study. Although the programme was delivered on a universal whole-school basis, if delivered on a more targeted basis, in a staged fashion and matched more appropriately to the levels of the students it may provide improved outcomes.

Future research and publications

Academic publications reporting the results of the trial are currently in preparation. The evaluation team hope to bring them to publication in the near future.

References

Ainscow et al, (2012) An Evaluation of The Communication Trust's 'Talk of the Town' Project. Centre for Equity in Education, University of Manchester: Manchester, UK.

Barber, B. K. & Olsen, J. A. (2004). Assessing the transitions to middle and high school. *Journal of Adolescent Research,* 19(1), 3–30.

Benner, G. J. (2005). Language skills of elementary-aged children with emotional and behavioral disorders*. Great Plains Research*, 15(2), 251–265.

Benner, G. J, Nelson, J. R. & Epstein, M. H. (2002). The language skills of children with emotional and behavioural disorders: A review of the literature. *Journal of Emotional and Behavioural Disorders,* 10, 43–59.

Bishop, D. V. M. & Adams, C. (1990). A prospective study of the relationship between specific language impairment, phonological disorders and reading retardation. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 31(7), 1027–1050.

Blischak, D. M., Shah, S. D., Lombardino, L. J. & Chiarella, K. (2004). Effects of phonemic awareness instruction on the encoding skills of children with severe speech impairment. *Disability and Rehabilitation*, 26(21–22), 1295–1304.

Borman, G. D., Slavin, R. E., Cheung, A., Chamberlain, A., Madden, N. A. & Chambers, B. (2007). Final reading outcomes of the national randomized field trial of Success for All. *American Educational Research Journal,* 44(3), 701–731.

Botting, N. (2005). Non-verbal cognitive development and language impairment. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, 46(3), 317–26.

Boyle J., Gillham B. & Smith N. (1996). Screening for early language delay in the 18–36-month age range: the predictive validity of tests of production and implications for practice. *Child Language Teaching and Therapy*, 12, 113–127.

Bronson, C. E. (2010). *The mosaic of small high school reform: A case study of one urban high school.* Dissertation Abstracts International Section A: Humanities and Social Sciences. 70(11-A), 4127.

Burge, B., Styles, B., Brzyska, B., Cooper, L., Shamsan, Y., Saltini, F. & Twist, L. (2010). *New Group Reading Test (NGRT).* Third Edition. London: GL Assessment.

Cirrin, F. M. & Gillam, R. B. (2008). Language intervention practices for school-age children with spoken language disorders: A systematic review. *Language, Speech, and Hearing Services in Schools*, 39(1), S110–S137.

Cockerill, M. & Thurston, A. (2015). Improving fidelity to treatment during randomised controlled trials in schools by engaging teachers in the design process during a developmental study. Chapter 12. In R. Gilles, Collaborative Learning: Developments in Research and Practice. New York, NY: Nova Science Publishing.

Collins, L. M., Schafer, J. L., & Kam, C-M. (2001). A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychological Methods*, 6, 330–351.

Conti-Ramsden, G., Botting, N. & Faragher, B. (2001). Psycholinguistic markers for specific language impairment (SLI). *Journal of Child Psychology and Psychiatry*, 42(6), 741–748.

Crowe, L. K. (2003). Comparison of two reading feedback strategies in improving the oral and written language performance of children with language-learning disabilities. *American Journal of Speech-Language Pathology*, 12(1), 16–27.

Department for Education (2015). National curriculum assessments at key stage 2 in England, 2015 (revised). Retrieved from: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/483897/SFR47_2015_text.pdf (15th January 2015)

Dockrell, J. & King, D. (2007). Talking time. Institute of Education: London.

Dockrell, J. E. & Lindsay, G. (2001). Children with Specific Speech and Language Difficulties—the teachers' perspective. *Oxford Review of Education*, *27*(3), 369–394.

Donlan, C., Cowan, R., Newton, E. J. & Lloyd, D. (2007). The role of language in mathematical development: Evidence from children with specific language impairments. *Cognition*, 103(1), 23–33.

Education Analytical Services (2008). Evaluation of Free School Meal trial for P1 to P3 pupils. Education Analytical Services, Scottish Government, UK. Retrieved from: http://dera.ioe.ac.uk/9515/1/0064986.pdf (27th October 2015)

Enders, C. K. (2010). *Applied missing data analysis.* New York, NY: The Guilford Press.

Finger, J. A. & Silverman, M. (1966). Changes in academic performance in the junior high school. *Personal and Guidance Journal*, 45, 157–164.

Gillon, G. T. (2000). The efficacy of phonological awareness intervention for children with spoken language impairment. *Language, Speech, and Hearing Services in Schools*, 31(2), 126–141.

Goswami, U. C., & Bryant, P. (1990). *Phonological skills and learning to read*. Psychology Press.

Hart, B. & Risley, T.R. (1995). Meaningful differences in the everyday experiences of young American children. Baltimore: Paul Brooke Publishing Co.

Hedges, L. & Hedberg, E. (2007). Intraclass correlation values for planning group-randomized trials in education*. Educational Evaluation and Policy Analysis,* 29(1), 60–87.

Holford, A. (2012). Take-up of Free School Meals: Price effects and peer effects. Institute for Social and Economic Research, University of Essex, UK. Retrieved from: https://www.iser.essex.ac.uk/research/publications/working-papers/iser/2012-12.pdf (27th October 2015)

Law, J., Garrett, Z. & Nye, C. (2004). The Efficacy of Treatment for Children with Developmental Speech and Language Delay/Disorder, A Meta-Analysis. .*Journal of Speech, Language, and Hearing Research*, 47(4), 924–943.

Law, J., Lee, W., Roulstone, S., Wren, Y., Zeng, B. & Lindsay, G. (2012). What Works: Interventions for children and young people with speech, language and communication needs. *Department of Education Research Report*, *DFE-RR247*(BCRP10), 36.

Law, J., Rush, R., Parsons, S. & Schoon, I. (2009). Modeling developmental language difficulties from school entry into adulthood: results from the BCS70 birth cohort literacy,

mental health and employment outcomes. *Journal of Speech, Language and Hearing Research*, 52, 1401–1416.

Law, J. & Sivyer, S. (2003). Promoting the communication skills of primary school children excluded from school or at risk of exclusion: An intervention study. Child Language Teaching and Therapy, 19(1), 1–25.

Nathan, L., Stackhouse, J., Goulandris, N. & Snowling, M. J. (2004). Educational consequences of developmental speech disorder: Key Stage 1 National Curriculum assessment results in English and mathematics. *The British Journal of Educational Psychology*, 74(Pt 2), 173–186.

National Equality Panel (2010). An anatomy of economic inequality in the UK. London: Government Equalities Office.

Peers, I.S., Lloyd, P. & Foster, C. (1999) British standardisation of the CELF. *The Psychological Corporation's Speech and Language Assessment.* London: Psychological Corporation.

Petersen, A. C. & Crockett, L. (1985). Pubertal timing and grade effects on adjustment. *Journal of Youth and Adolescence*, 14, 191–206.

Puma, M., Bell, S., Olsen, R. & Price, C. (2009). *Missing data issues in randomized control trials: What to do when data are missing.* Washington, DC: Institute of Education Sciences, U.S. Department of Education.

Raudenbush, S. W. & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods.* 2nd Ed. Newbury Park, CA: Sage Publications.

Ravitz, J. (2010). Beyond changing culture in small high schools: Reform models and changing instruction with project-based learning. *Peabody Journal of Education*, 85, 290–312.

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys.* Hoboken, NJ: Wiley.

Sadler, J. (2005). Knowledge, attitudes and beliefs of the mainstream teachers of children with a preschool diagnosis of speech/language impairment. *Child Language Teaching and Therapy*, 21(2), 147–164.

Schafer, J. L. (1997). Analysis of incomplete multivariate data. Boca Raton, FL: Chapman & Hall.

Schafer, J. L. & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, 7, 147–177.

Segers, E. & Verhoeven, L. (2005). Effects of lengthening the speech signal on auditory word discrimination in kindergartners with SLI. *Journal of Communication Disorders*, 38(6), 499–514.

Snowling, M. J., Adams, J. W., Bishop, D. V. M. & Stothard, S. E. (2001). Educational attainments of school leavers with a preschool history of speech-language impairments. *International Journal of Language & Communication Disorders*, 36(2), 173–183.

The Poverty Site (2015). Educational attainment at age 16. Retrieved from: http://www.poverty.org.uk/26/index.shtml (14th November 2015).

Throneburg, R. N., Calvert, L. K., Sturm, J. J., Paramboukas, A. A. & Paul, P. J. (2000). A Comparison of Service Delivery Models. Effects on Curricular Vocabulary Skills in the School Setting. *American Journal of Speech-Language Pathology*, 9(1), 10–20.

Tymms, P., Merrell, C., Thurston, A., Andor, J., Topping, K. & Miller, D. (2011). Improving attainment across a whole district: school reform through peer tutoring in a randomized controlled trial. *School Effectiveness and School Improvement: An International Journal of Research, Policy and Practice*, 22(3), 265–289.

Van Kleeck, A., Gillam, R. B. & McFadden, T. U. (1998). A study of classroom-based phonological awareness training for preschoolers with speech and/or language disorders. *American Journal of Speech-Language Pathology*, 7(3), 65–76.

Walker, D., Greenwood, C., Hart, B. & Carta, J. (1994). Prediction of school outcomes based on early language production and socioeconomic factors. *Child Development*, 65, 606–621.

Wellington, W. & Wellington, J. (2002). Children with communication difficulties in mainstream science classrooms. *School Science Review*, 83(305), 81–92.

West, B. T., Welch, K. B. & Galecki, A. T. (2014). *Linear mixed models: A practical guide using statistical software*. 2nd ed. Boca Raton, FL: Chapman and Hall/CRC Press.

Appendix A

Parental Information Form

Your child has been invited to take part in a research study. Before you decide whether you would like them to take part, it is important for you to understand why the research is being conducted and what it will involve. Please take time to read the following information carefully. As you may know your school has agreed to be part of a programme called Talk of the Town. The programme has been in action in your child's school for two years. It is a partnership between The Communication Trust and your child's school. The programme is being evaluated by a team from the Centre for Effective Education at Queen's University Belfast, Northern Ireland.

We would like your permission for your child to take part in the evaluation. Participation will involve our researcher administering a child-friendly test to establish your child's reading and language skills. Completion of the test will last approx. 20 minutes and take place during normal classroom contact time. All information gathered from the test will remain confidential and used for evaluation purposes only. We are evaluating whether the programme has benefited your child (compared to children who have not had the programme), not how well your child is doing in school. No children or schools will be identified by name in any reports as all reports will only identify those who had, and did not have the programme. Data will be held securely for a minimum of five years by Queen's University Belfast and then securely destroyed. Electronic data will be held at a secure site on an encrypted computer.

As participation is voluntary your child is free to withdraw from the study at any time up until the point that any reports derived from the project are published. Any reports will not mention individual names nor even the name of schools that participated in the project. A decision to participate (or not) will not affect your relationship with the school or with Queen's University Belfast. The project has ethical approval from Queen's University Belfast School of Educational ethics Committee.

If you are willing to allow your child to participate please complete the details on the consent form, sign it, and return to your child's class teacher.

If you have any questions please don't hesitate to contact Jenny Davison, 6 College Green, School of Education, Queen's University Belfast, Belfast or phone on 028 9097 5294 or via email at j.davison@qub.ac.uk.

If you have a complaint regarding the work then contact Allen Thurston at Queen's University Belfast on 07964916186 or via a.thurston@qub.ac.uk

Parental Consent Form

Please sign and return this form to your child's class teacher as soon as possible if you are willing for your child to take part in the study.

Consenting to participate in the research project:

Child's Name ……..……

I give permission for my child (name as above) to take part in this research activity as part of the Talk of the Town research project.

I understand what data the team will collect.

I understand how data will be stored and used.

I understand that I am free to withdraw consent at any time up until the point that data is published.

YES ☐ **NO** ☐

Name ……..

Signed ……..……

Parent / Guardian (delete as appropriate)

Date ……..……

Talk of the Town programme is provided by the Communication Trust and is funded by The Education Endowment Foundation: www.thecommunicationtrust.org.uk/projects/talk-of-the-town

Parental Information Form

Dear Parent/Guardian,

Your child has been invited to continue to take part in a research study. Before you decide whether you would like them to continue to take part, we would like to inform you about why the next stage of research is being conducted and what it will involve. We would appreciate it if you could please take a few moments to read the following information carefully.

Your child's school has been involved in a program called *Talk of the Town* for two years. It is a partnership between The Communication Trust and your child's school. The programme is being evaluated by a team from the Centre for Effective Education at Queen's University Belfast, Northern Ireland.

With your consent, your child took part in an evaluation of their reading and language skills last year. We now require access to the Key Stage 1 and Key Stage 1 Phonics scores from the National Pupil Database for the children who already took part in the evaluation. **Therefore, this letter and consent form is asking you to tell us if you do NOT allow us to access Key Stage scores from the National Pupil Database.**

We require these scores to establish how successful the Talk of the Town program has been in improving children's educational attainment. We are evaluating whether the programme has benefited children (compared to children who have not had the programme), and not how well your child is doing in school. Data will be held securely for a minimum of five years by Queen's University Belfast and then securely destroyed. Electronic data will be held at a secure site on an encrypted computer.

As participation is voluntary, your child is free to withdraw from the study at any time up until the point that any reports derived from the project are published. Any reports will not mention individual names or even the name of schools that participated in the project. A decision to participate (or not) will not affect your relationship with your child's school or with Queen's University Belfast. The project has ethical approval from Queen's University Belfast School of Educational ethics Committee.

If you **do NOT wish** us to access your child's Key Stage data from the National Pupil Database, please complete the details on the consent form below, sign it, and return to your child's class teacher. If you are happy for us to access this data no further action from you is required.

If you have any questions please don't hesitate to contact Patrick Stark, 6 College Green, School of Education, Queen's University Belfast, Belfast or phone on 028 9097 5294 or via email at p.stark@qub.ac.uk.

If you have a complaint regarding the work then contact Allen Thurston at Queen's University Belfast on 07964916186 or via a.thurston@qub.ac.uk

Parental Consent Form

Please return this form to your child's class teacher if you **are NOT** willing for your child's Key Stage 1 scores to be accessed from the National Pupil Database.

I **do NOT** give permission for my child's Key Stage 1 scores to be accessed in the National Pupil Database as part of the Talk of the Town research project.

Your Child's Name ……..……

Your Name ……..

Your Signature ……..……

Parent / Guardian (delete as appropriate)

Date ……..……

Talk of the Town programme is provided by the Communication Trust and is funded by The Education Endowment Foundation: www.thecommunicationtrust.org.uk/projects/talk-of-the-town

Appendix B

Table A1: NGRT Analyses: HLM Results for Observed and Multiply Imputed Data

		Unconditional		ANCOVA		Student covariates	
Parameter		**Observed**	**MI**	**Observed**	**MI**	**Observed**	**MI**
Fixed effects							
Student-level	Intercept (γ_{00})	99.69*** (0.50)	99.43*** (0.49)	99.55*** (0.68)	99.83*** (0.67)	100.96*** (0.74)	101.54*** (0.73)
	NGRT pretest (γ_{01})			9.29*** (0.18)	9.31*** (0.19)	8.50*** (0.20)	8.51*** (0.20)
	Gender (γ_{02})					-0.50 (0.37)	-0.60 (0.37)
	Year (γ_{03})					0.47 (0.37)	0.24 (0.37)
	Minority (γ_{04})					-0.38 (0.75)	-0.57 (0.74)
	Free school meals (FSM) (γ_{05})					-1.03* (0.40)	-1.24** (0.41)
	English as additional language (EAL) (γ_{06})					0.81 (0.93)	0.53 (0.96)
	Special education (SEN) (γ_{07})					-4.19*** (0.48)	-4.19*** (0.48)
School-level	Talk of Town (γ_{10})			-0.46 (1.00)	-0.81 (0.98)	-0.43 (0.97)	-0.76 (0.96)
Random effects	Between-school (τ_{00})	11.67*** (2.88)		13.13*** (2.84)		12.40*** (2.70)	
	Within-school (σ^2)	180.77*** (4.98)					

	Control (σ^2)		77.15***	75.32***
			(2.89)	(2.83)
	Talk of Town (σ^2)		118.15***	113.66***
			(4.80)	(4.63)
Model fit	AIC	21,743.5	20,014.1	19,870.7

Note. NGRT pretest scores were group-centered within schools; standard errors (SE) follow parameter estimates in parentheses; MI = Multiple imputation; AIC = Akaike's Information Criterion; $^{***}p < .001$; $^{**}p < .01$; $^{*}p < .05$.

Table A2: FSM Subgroup Analyses of Post-test NGRT Scores: HLM Results for Observed and Multiply Imputed Data

		Unconditional		ANCOVA		Student covariates	
Parameter		**Observed**	**MI**	**Observed**	**MI**	**Observed**	**MI**
Fixed effects							
Student-level	Intercept (γ_{00})	96.91*** (0.63)	96.30*** (0.59)	98.80*** (0.77)	98.71*** (0.73)	100.49*** (0.89)	100.38*** (0.80)
	NGRT pretest (γ_{01})			8.34*** (0.32)	8.30*** (0.32)	7.24*** (0.35)	7.18*** (0.35)
	Gender (γ_{02})					-0.27 (0.64)	-0.20 (0.66)
	Year (γ_{03})					0.37 (0.62)	0.46 (0.62)
	Minority (γ_{04})					0.08 (1.20)	0.35 (1.19)
	English as additional language (EAL) (γ_{05})					-1.41 (1.62)	-1.54 (1.61)
	Special education (SEN) (γ_{06})					-5.74*** (0.77)	-5.84*** (0.81)
School-level	Talk of Town (γ_{10})			-1.17 (1.17)	-1.45 (1.11)	-1.33 (1.14)	-1.54 (1.10)
Random effects	Between-school (τ_{00})	13.51*** (4.28)		13.65*** (3.66)		13.02*** (3.54)	
	Within-school (σ^2)	166.45*** (7.73)					
	Control (σ^2)			78.55*** (4.96)		74.51*** (4.74)	
	Talk of Town (σ^2)			125.48*** (8.61)		119.11*** (8.20)	
Model fit	AIC	7,848.8		7,341.0		7,270.0	

Note. NGRT pretest scores were group-centered within schools; standard errors (SE) follow parameter estimates in parentheses; MI = Multiple imputation; AIC = Akaike's Information Criterion; $^{***}p < .001$; $^{**}p < .01$; $^{*}p < .05$.

Table A3: Subgroup Analyses of Post-test NGRT Scores Predicted by ACE Language Scores: HLM Results for Observed and Multiply Imputed Data

		Unconditional		ANCOVA		Student covariates	
Parameter		**Observed**	**MI**	**Observed**	**MI**	**Observed**	**MI**
Fixed effects							
Student-level	Intercept (γ_{00})	92.44*** (0.76)	92.42*** (0.71)	92.80*** (0.97)	92.75*** (0.93)	97.77*** (3.05)	98.44*** (3.10)
	NGRT pretest (γ_{01})	3.49*** (0.62)	3.51*** (0.65)	3.24*** (0.59)	3.31*** (0.62)	2.80*** (0.60)	2.90*** (0.62)
	ACE pretest (γ_{01})	3.12*** (0.61)	3.29*** (0.60)	1.53*** (0.59)	3.26*** (0.58)	3.02*** (0.60)	3.17*** (0.59)
	Gender (γ_{02})					-1.67 (1.11)	-1.76 (1.08)
	Minority (γ_{03})					-3.30 (2.76)	-3.96 (2.84)
	Free school meals (FSM) (γ_{04})					0.22 (1.16)	-0.02 (1.15)
	English as additional language (EAL) (γ_{05})					1.98 (2.55)	2.38 (2.60)
	Special education (SEN) (γ_{06})					-3.58** (1.26)	-3.63** (1.24)
School-level	Talk of Town (γ_{10})			-0.70 (1.53)	-0.65 (1.45)	-0.89 (1.48)	-1.02 (1.43)
Random effects	Between-school (τ_{00})	16.82** (6.71)		16.91** (6.56)		14.85** (6.29)	
	Within-school (σ^2)	90.19*** (8.30)					
	Control (σ^2)			64.61*** (8.17)		65.98*** (8.39)	
	Talk of Town (σ^2)			118.37*** (15.34)		109.57*** (14.67)	
Model fit	AIC	2,235.5		2,221.3		2,191.6	

Note. NGRT pretest scores were group-centered within schools; standard errors (SE) follow parameter estimates in parentheses; MI = Multiple imputation; AIC = Akaike's Information Criterion; $^{***}p < .001$; $^{**}p < .01$; $^{*}p < .05$; $^{\dagger}p < .10$.

Table A4: ACE Analyses: HLM Results for Observed and Multiply Imputed Data

		Unconditional		ANCOVA		Student covariates	
Parameter		**Observed**	**MI**	**Observed**	**MI**	**Observed**	**MI**
Fixed effects							
Student-level	Intercept (γ_{00})	91.69*** (0.93)	91.41*** (0.87)	90.98*** (1.26)	91.12*** (1.21)	54.76*** (11.17)	53.64*** (12.14)
	NGRT pretest (γ_{01})	7.88*** (0.61)	7.67*** (0.65)	7.88*** (0.60)	7.67*** (0.65)	7.34*** (0.61)	7.13*** (0.66)
	Gender (γ_{02})					0.45 (1.16)	0.54 (1.38)
	Minority (γ_{03})					1.19 (2.95)	0.50 (3.49)
	Reading Age (γ_{04})					0.46*** (0.13)	0.48** (0.14)
	Free school meals (FSM) (γ_{05})					0.87 (1.54)	0.84 (1.77)
	English as additional language (EAL) (γ_{06})					-2.02 (2.76)	-1.70 (3.01)
	Special education (SEN) (γ_{07})					-2.45† (1.39)	-1.90 (1.49)
School-level	Talk of Town (γ_{10})			1.42 (1.87)	0.58 (1.74)	3.92† (2.00)	3.17 (1.95)
	Talk of Town x FSM (γ_{20})					-6.59** (2.53)	-6.12* (2.65)
Random effects	Between-school (τ_{00})	34.24*** (9.97)		33.76*** (10.01)		26.39** (8.87)	
	Within-school (σ^2)	91.49*** (8.56)					
	Control (σ^2)			78.13*** (9.95)		75.10*** (9.64)	
	Talk of Town (σ^2)			106.57*** (14.44)		102.25*** (14.25)	
Model fit	AIC	2,205.7		2,199.1		2,157.0	

Note. NGRT pretest scores were group-centered within schools; standard errors (SE) follow parameter estimates in parentheses; MI = Multiple imputation; AIC = Akaike's Information Criterion; $^{***}p < .001$; $^{**}p < .01$; $^{*}p < .05$; $^{\dagger}p < .10$.

Table A5: FSM Subgroup Analyses of Post-test ACE Scores: HLM Results for Observed and Multiply Imputed Data

		Unconditional		ANCOVA		Student covariates	
Parameter		**Observed**	**MI**	**Observed**	**MI**	**Observed**	**MI**
Fixed effects							
Student-level	Intercept (γ_{00})	90.21*** (1.17)	89.63*** (1.21)	91.41*** (1.38)	91.24*** (1.45)	46.04*** (17.03)	46.89*** (19.05)
	NGRT pretest (γ_{01})	8.32*** (0.99)	8.38*** (1.10)	8.22*** (0.96)	8.37*** (1.09)	7.02*** (0.97)	7.24*** (1.08)
	Gender (γ_{02})					0.65 (1.95)	0.93 (2.32)
	Minority (γ_{03})					0.64 (6.13)	2.25 (6.50)
	Reading Age (γ_{04})					0.60** (0.20)	0.56* (0.22)
	English as additional language (EAL) (γ_{05})					-3.38 (7.75)	-0.27 (8.63)
	Special education (SEN) (γ_{06})					-4.39* (2.09)	-4.35† (2.27)
School-level	Talk of Town (γ_{10})			-3.09 (2.44)	-3.74 (2.43)	-3.41 (2.22)	-3.67 (2.30)
Random effects	Between-school (τ_{00})	23.40† (15.30)		20.59† (13.78)		8.32 (11.69)	
	Within-school (σ^2)	106.99*** (17.34)					
	Control (σ^2)			87.46*** (17.35)		84.40*** (16.78)	
	Talk of Town (σ^2)			141.07*** (33.00)		146.44*** (34.06)	
Model fit	AIC	954.2		946.3		914.9	

Note. NGRT pretest scores were group-centered within schools; standard errors (SE) follow parameter estimates in parentheses; MI = Multiple imputation; AIC = Akaike's Information Criterion; $^{***}p < .001$; $^{**}p < .01$; $^{*}p < .05$; $^{\dagger}p < .10$.

Appendix C

Linear mixed models (LMMs) were used to compare the relative effects of the *Talk of the Town* intervention and control conditions on the primary (reading) and secondary (language) outcomes while accounting for the clustered structure of the data.

Step 1 – Unconditional analyses. First, a fully unconditional model was used to assess whether there was significant variation in achievement across schools. In this model, the level 1 equation was written as:

$$Yij = \beta 0j + rij,$$

where Yij is the achievement score for student *i* at school *j*, and rij is the level-1 residual variance assumed ~ $N(0, \sigma^2)$. The unconditional model's level 2 equation was written as:

$$\beta 0j = \gamma 00 + u0j,$$

where β0j is the mean posttest score in school *j*, and μ0j is the random effect (i.e., between-school variance) for school *j* assumed to be ~ $N(0,\tau 00)$.

Step 2 – Conditional analyses. After determining whether there is significant variation between schools, the main analyses involved a standard intent to treat design and used a one-way ANCOVA model with random effects (i.e., two-level HLM) to assess the relationship between school-level treatment status (treatment versus control) and student achievement outcomes while controlling for prior attainment. Standardised student achievement outcomes were used for all analyses. The level 1 model for the one-way ANCOVA was written as:

$$Y_{ij} = \beta_{0j} + \beta_{1j}\,(\text{Pretest})_{ij} + r_{ij},$$

where:

- Y_{ij} is the achievement score for student *i* at school *j*;
- $(\text{Pretest})_{ij}$ is pretest achievement score for each student;
- β_{0j} is the mean posttest score in school *j*;
- β_{1j} is the difference between *i*th student's pretest score and the group-mean centered $(X_{ij} - \bar{X}_{.j})$ pretest score at school *j*;
- r_{ij} is the level-1 residual variance assumed $\sim N(0, \sigma^2)$.

At level 2, we estimated treatment effects on the mean posttest score in school *j*. The level 2 model was written as:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(\text{TalkTown})_j + u_{0j};$$

$$\beta_{1j} = \gamma_{10};$$

where:

- γ_{00} is the adjusted grand mean posttest score;

- $\gamma 01$ is the difference between the *Talk of the Town* schools and control school achievement scores, where TalkTown = 1 if school *j* was assigned to the *Talk of the Town* condition and 0 if not;

- $\mu 0j$ is the random effect (i.e., between-school variance) for school *j* associated with the adjusted mean posttest scores, assumed to be ~ $N(0,\tau_{|\text{Covariates}})$.

Step 3 – Student covariates. In Step 3, we added a vector of student characteristics to examine if *Talk of the Town* effects remain even after controlling for student characteristics such as gender, year in school, minority status, free school meals (FSM), English as additional language (EAL), and special education stage (SEN). The level 1 model used for the secondary analyses was written as:

$$Yij = \beta 0j + \beta 1j(\text{Pretest})ij + \beta 2j(\text{Gender})ij + \beta 3j(\text{Year})ij + \beta 4j(\text{Minority})ij + \beta 5j(\text{FSM})ij + \beta 6j(\text{EAL})ij + \beta 7j(\text{SEN})ij + rij$$

where

- Yij is the achievement score for student *i* at school *j*;

- $\beta 0j$ is the adjusted mean posttest score of Year 2, control-group female students who are not a minority, receive free school meals, learning English as an additional languague or enrolled in special education;

- $\beta 1j$ is the difference between *i*th student's pretest score and the group-mean centered ($X_{ij} - \bar{X}_{\cdot j}$) pretest score at school *j*;

- (Pretest)ij is pretest achievement score for each student;

- $\beta 1j - \beta 7j$ are the student-level dummy variables indicating differences in achievement scores associated with Gender (1 = male), Year (1 = Year 5), minority status (1 = minority), free school meals (FSM; 1 = yes), English as an additional language (EAL; 1 = yes), and special education status (SEN; 1 = yes), respectively;
- rij is the level-1 residual variance assumed ~ $N(0, \sigma^2)$.

The level 2 model used for the secondary analyses was written as:

$$\beta 0j = \gamma 00 + \gamma 01(\text{TalkTown})j + u0j;$$

$$\beta kj = \gamma k0 \text{ for } k = 1...7;$$

$$\beta_{kj} = \gamma_{k0} \quad for\ k = 1 \ldots 7$$

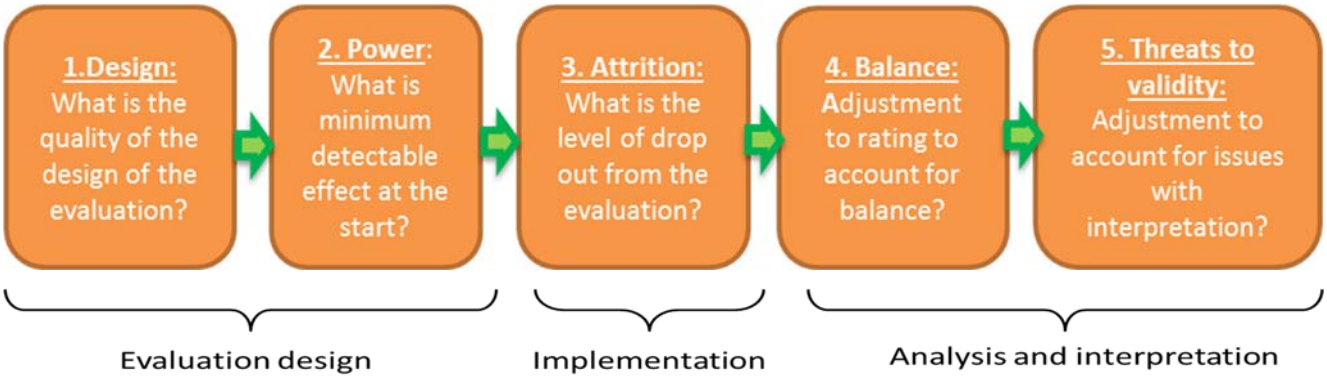
where:

- $\gamma 00$ is the adjusted grand mean posttest score;
- $\gamma 01$ is the difference between the *Talk of the Town* schools and control school achievement scores, where TalkTown = 1 if school *j* was assigned to the *Talk of the Town* condition and 0 if not;
- $\gamma 10 - \gamma 70$ are the fixed effects representing the difference in posttest scores associated with student characteristics;
- $\mu 0j$ is the random effect (i.e., between-school variance) for school *j* associated with the adjusted mean posttest scores, assumed to be ~ $N(0,\tau_{|\text{Covariates}})$.

Step 4 – Subgroup analyes. In Step 4, we looked at the predictive association between language (as measured by ACE) and posttest reading scores (as measured by NGRT) by repeating Steps 1-3 with the subgroup of students with pretest ACE scores.

Step 5 – FSM subgroup analyes. Finally, and in keeping with the EEF's analysis policies, we repeated Steps 1-3 with the free school meals (FSM) subgroup. Although this diverges from the protocol, these additional analyses that were not mentioned in the protocol were undertaken to take account fo revised guidance and advice provided by EEF.

Appendix D: Padlock rating



Rating	1. Design	2. Power (MDES)	3. Attrition	4. Balance	5. Threats to validity
5 🔒	Fair and clear experimental design (RCT)	< 0.2	< 10%	Well-balanced on observables	No threats to validity
4 🔒	Fair and clear experimental design (RCT, RDD)	< 0.3	< 20%		
3 🔒	Well-matched comparison (quasi-experiment)	< 0.4	< 30%		
2 🔒	Matched comparison (quasi-experiment)	< 0.5	< 40%		
1 🔒	Comparison group with poor or no matching	< 0.6	< 50%		
0 🔒	No comparator	> 0.6	> 50%	Imbalanced on observables	Significant threats

The final security rating for this trial is 4 🔒. This means the findings are secure.

The trial was designed as a well-powered, cluster randomised effectiveness trial. The sample size was able to detect a MDES of less than 0.1 at analysis. Attrition was low at the cluster level. Although individual attrition is secondary, this was an issue. The sample was balanced at baseline. There is a minor potential threat to validity from the lack of information about control schools (which was not part of the evaluation protocol).

.Therefore, the overall padlock rating is 4 🔒.

Appendix E: Cost rating

Cost ratings are based on the approximate cost per pupil of implementing the intervention over one year. Cost ratings are awarded using the following criteria.

Cost	Description
£	*Very low:* less than £80 per pupil per year.
£ £	*Low:* up to about £170 per pupil per year.
£ £ £	*Moderate:* up to about £700 per pupil per year.
£ £ £ £	*High:* up to £1,200 per pupil per year.
£ £ £ £ £	*Very high:* over £1,200 per pupil per year.

You may re-use this document/publication (not including logos) free of charge in any format or medium, under the terms of the Open Government Licence v2.0.

To view this licence, visit www.nationalarchives.gov.uk/doc/open-government-licence/version/2 or email: psi@nationalarchives.gsi.gov.uk

Where we have identified any third party copyright information you will need to obtain permission from the copyright holders concerned. The views expressed in this report are the authors' and do not necessarily reflect those of the Department for Education.

This document is available for download at www.educationendowmentfoundation.org.uk

Education
Endowment
Foundation

The Education Endowment Foundation
9th Floor, Millbank Tower
21–24 Millbank
London
SW1P 4QP
www.educationendowmentfoundation.org.uk