

Evaluation Summary	
Age range	Primary (Year 2, 3, 5 and 6)
Number of pupils	3720
Number of schools	62
Design	Cluster randomised controlled trial, with randomisation at the school level
Primary Outcome	Reading

## Overview

Talk of the Town is a Toolkit of five language-based interventions. Two important points should be taken into account when selecting the most appropriate evaluation methods:

- In evaluating Talk of the Town it would be important to conceptualise the Toolkit of five interventions as one intervention. In reality the Toolkit may result in a number of separate interventions, dependent on the individual needs of students and interventions and training selected by schools. The evaluation will focus on the overall impact on professional practices and pedagogies of teachers and the impact it has on student learning in reading.
- The link between language-based interventions and enhanced literacy will need to be established as an secondary outcome for the project.

Therefore the overall research question and associated secondary questions will be as follows:

1. What are the effects on student reading levels in Granada Learning (GL) New Group Reading Test (NGRT) of use of the 'Talk of the Town'?
2. Secondary research questions:
  - a. What are the effects on student oral language levels as measured by GL Assessment of Comprehension and Expression 6-11 (ACE) of use of the 'Talk of the Town'?
  - b. How do language levels predict reading outcomes?
  - c. How do teachers, members of school senior management team, speech and language therapists and local authority managers perceive that delivery of the implementation can implementation be optimised in future full-scale initiatives?

## Outcome measures

- The primary outcome measure of the study will be reading as measured by the New Group Reading Test for Year 2/3 and Year 5/6 students.
- Secondary outcome measure will be language of Year 2/3 students as measured by GL ACE assessment.
- Tertiary outcomes measures will be KS results 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.

## Sample

*Effect Sizes* from a previous study undertaken by The Communications Trust have been used to calculate sample sizes for HLM statistical analysis. *Effect Sizes* on previous studies varied widely. The

delivery team reported that maximum *Effect Sizes* were reported when the design protocol for Talk of the Town were adhered to. In the design experiment reported *Effect Sizes* on standardised scores were:

GL ACE Test (A language based test offering a measure of competence in spoken language)

1. Naming section=*ES* max 0.8, min 0.22, mean *ES*=0.51
2. Sentence completion= *ES* max 0.82, min 0.35, mean *ES*=0.59

BUS Story test (A standardised test of reading)

1. *ES* max 0.23, min 0.15, mean *ES*=0.19

If we assume that the intervention could have an *Effect Size* of 0.25, then we would require about 62 schools to participate, 31 intervention, 31 control to assess outcomes in academic areas. This would probably be a pragmatic sample size. It is most likely that this figure would be enough to detect effects in the ACE (i.e. in use of language) test, but may be underpowered to detect effects in reading tests (i.e. in literacy). It is proposed that reading measures are taken from two-year groups in the school (one from Year 2 and one from Year 5) and ACE language tests are conducted in one year group to establish a baseline for longitudinal study and to model how language effects transmit to effects in literacy during the two-year project and in any subsequent follow-up study. These predictions should hold true for the trial. The *Effect Sizes* were calculated on standardised scores for each study. This means that although not a randomised design, increases and effects were against the wider 'standardisation population'.

### **Recruitment**

It may be desirable to work across at least two/three local authority areas to generate this sample. Strong relationships with local authorities selected would be key to successful recruitment. In addition the recruited schools should have a 'memorandum of understanding' so that they are away of their responsibilities and rights in respect of the project. Recruitment will be undertaken and managed by Communications Trust (CT). Allen Thurston will support CT in the recruitment process. He will provide input on the experimental design and measures to be used.

### **Tests, Assessments & Measures**

It is proposed that two outcome measures are used in this evaluation:

***GL New Group Reading Test (NGRT):*** This test of reading ability will be administered at pre and post-test in control and intervention schools with one class of Year 2 (post-test at the end of Year 3) and one class of Year 5 (post-test at the end of Year 6) students. Pre-test NGRT would be implemented by teachers, but coded and processed by QUB. Post-test NGRT would be implemented and coded by staffs from QUB blind to condition.

***GL Assessment of Comprehension and Expression 6-11:*** This test of language will be undertaken with a randomly selected sub-set of 6 students from the bottom half of the class reading attainment profile (as measured by NGRT pre-test) pre and post-test from each school class in Year 2. The test will be repeated with the same students at the end of Year 3 (note it will not be possible to test all students in the sample due to the high cost of this test administration). In terms of the ACE assessment the following sub-sections will be assessed:

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

Other outcome measures:

**Key Stage (KS) data:** Mean KS1 test scores for current Year 5 students in the quantitative measures group will be used to assign schools to condition pairwise as indicated below. KS1 data will be collected from students at the end of Year 2 to provide an interim report on differences between profiles in implementation and control schools. KS2 test data will be collected in language and mathematics for both intervention and control groups at the end of Year 6 for students who started the project in Year 5.

**Other data from National Pupils Data Base (NPDB):** Where possible data will be collected from NPDB on the Year 1 phonics screening check assessment at pre-test and the Year 6 Spelling, Punctuation and Grammar (SPAG) test for Year 6 at post-test. These data do not form primary or secondary outcome measures, but may be useful in interrogating links between language and literacy development.

**Process data and evaluations focussing on how to optimise the intervention in follow-up projects:**

**Interviews** 20 semi-structured interviews will be 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) to explore their perceptions of what worked well in the project and how the intervention could be enhanced. Similar interviews will also be undertaken with one member of the SMT from each of these 20 schools and from a senior manager from each local authority that schools were situated within. Care will be taken to devise a method of undertaking interviews that does not contaminate knowledge of condition for other testing procedures.

**Questionnaire** A postal questionnaire will be issued to all teachers who have participated in the project as a Year 2, 3 5 or 6 teacher in a class that formed part of the quantitative data set. Teacher perceptions of the potential benefits of the interventions will be collected through a structured questionnaire. In addition teachers will be asked about wider impact on their students.

**Analysis:** The final data set will be collated by QUB. The data set will be analysed by a visiting senior research fellow to QUB (Dr Cary Roseth, QUB & Michigan State University) blind to condition. This will be done by removing the descriptor of variables in the data set prior to analysis and randomly assigning numbers to represent control and intervention conditions.

The instruments and measures are summarised in the table below:

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 (now at the end of Year 3)	Year 1 phonics assessment
New Group Reading Test	Year 2 Year 5	Year 2 (now end of Year 3) Year 5 (now end of Year 6)	Year 6 SPAG test data
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 Year6)
Teacher interviews and questionnaires		Interviews: 20 teachers from different schools who have taught in classes forming part of the quantitative data set Questionnaires: All teachers who have participated in the project as a Year 2, 3 5 or 6 teacher in a class that formed part of the quantitative data set.	
Senior management team (SMT) interviews		One member of SMT from each intervention school	
Local authority manager		One senior manager from each local authority involved in the project	

### Assignment to condition

We propose a design that has Talk of the Town as the intervention and a control condition with ‘Treatment as Usual’ that we are assured has equal access to Educational Psychology, Speech and Language Therapy Services and Health Support should they chose to make use of them (further inclusion criteria are included below). In this respect it would be treatment group (intervention) versus treatment as usual (control). Assignment to condition would be by block randomization. Block randomization would take 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 will be 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 assigned one each to intervention/control, schools three and four on the list similarly assigned and so on until each school is assigned. Block randomization is preferable as it minimises differences in gains for groups on the basis of prior attainment. Pre-test differences for Year 2 & Year 5 classes in New Group Reading Test would be analysed and adjusted for in the final model. Block randomization is still required as the ‘rate of gain’ may differ between classes having different starting points (hence the often quoted issue of a widening gap between high and low attaining students in school).

It is proposed that recruitment continues even when 62 schools are recruited. The first 62 schools recruited will be assigned to implementation or control condition. Any additional schools will be placed on a waiting list and also assigned to implementation or control condition. If a school drops out before the memorandum of agreement is signed, then the first school from either implementation or control condition will replace them in the study.

## Inclusion criteria

Primary schools that meet the following criteria will be eligible for inclusion in the study:

1. At least 20 Year 2 students in single form entry class and 20 Year 5 students in single form entry class.
2. Willing to be randomly assigned to condition at the school level.
3. Meet criteria acceptable to EEF which shall be that the majority of schools in the sample will be from a disadvantaged area.
4. Be willing to supply access to socio-economic and demographic data from students.
5. Access to Educational Psychology, Speech and Language Therapy Services and Health Support.
6. Access to KS1 data for students starting Year 5 in September 2013.
7. Willing to engage with 'talk of the Town' intervention and implement this in at least Year 2 and Year 5.

## Proposed analysis

A school-based intervention would allow hierarchical multiple regression analysis to be undertaken.

Built into this evaluation bid is time to fine-tune the equations that will be used for the HLM analysis. These will be defined before the study takes place. The following represents a draft proposal of what these equations may look like. The study will take place in  $j=62$  sites. Schools will have about  $n=30$  students per year group (although the figure could drop considerably lower ( $<20$ ) without influencing the integrity of data). The main analyses of Research Question A will involve a standard intent to treat design with a two-level HLM model to assess the relationship between school-level treatment status (treatment versus control), and student achievement outcomes. Student achievement outcomes will be used for analysis as Standardised scores. For the hierarchical analysis, the level 1 model is written as

$$Y_{ij} = \beta_{0j} + \beta_{1j}(\text{Pretest})_{ij} + \beta_{2j}(\text{Year})_{ij} + \beta_{3j}(\text{Gender}) + \beta_{4j}(\text{Ethnicity}) + r_{ij}$$

which represents the post-test achievement for student  $i$  in school  $j$  regressed on four covariates: pre-test, year, gender, and ethnicity. The term  $r_{ij}$  is the level-1 residual variance that remains unexplained after accounting for the covariates. Pre-test scores will be utilised to establish pre-intervention equivalence in control and intervention samples.

At level 2 of the model, we estimate treatment effects of the treatment on the mean post-test achievement outcome in school  $j$ . The fully specified level 2 model is written as

$$\begin{aligned}\beta_{0j} &= \gamma_{00} + \gamma_{01}(\text{Pre-test}) + \gamma_{02}(\text{Year}) + \gamma_{03}(\text{Gender}) + \gamma_{04}(\text{Ethnicity}) + \gamma_{05}(\text{SSL}) + u_{0j} \\ \beta_{1j} &= \gamma_{10} + \gamma_{11}(\text{SSL})_{j} + u_{1j} \\ \beta_{2j} &= \gamma_{20}\end{aligned}$$

where the mean post-test intercept for school  $j$ ,  $\beta_{0j}$ , is regressed on the school-level means for the covariates and the treatment indicator, plus a residual,  $u_{0j}$ . The pre-test/post-test slope,  $\beta_{1j}$ , is predicted by the treatment indicator and a school-specific residual, and because we have no substantive or theoretical reason to model the year, gender, or ethnicity slopes, are treated as fixed effects, predicted only by the school-level mean. If attrition of schools is low, as we expect, we will use case deletion, but if not, we will use multiple stochastic imputation for missing outcome data (see Puma et al., 2009).

The ability to detect a treatment effect at a certain level of power depends on several factors: intraclass correlation ( $\rho$ ), the correlation between covariate and post-tests ( $r$ ), the average number of students in each school ( $n$ ), and the number of school-level clusters ( $j$ ). In our power analysis, we used an estimate of the intra-class correlation,  $\rho$ , of 0.10 that represents the proportion of variance in achievement scores that we expect to find between schools. This estimate is derived from the range from .05 to .12 for low-achieving schools reported by Hedges & Hedberg (2007), and is

consistent with ICCs from previous studies (e.g., Borman et al., 2007). To increase statistical power, we also included four covariates. Based on data from previous studies, the proportion of variance in post scores explained by the covariates is estimated at  $r^2=0.60$ .

According to Optimal Design software, using estimates of  $\rho=0.10$ , and  $r^2=0.60$  and  $n=30$ , with an expected total sample for each pairwise comparison of 62 schools, treatment effects of  $ES=+0.25$  will be detected at an alpha level of  $p<.05$  with a power of .80, using a two-level HLM model with school as a random effect.

In addition to the main achievement analyses, similar analyses will be carried out to determine interactions between treatment and pre-test, ethnicity, and gender. There will be adequate power for each of these analyses, as the numbers of schools will be identical to those for the main analyses, and power for CRTs is primarily driven by the number of school-level clusters  $j$  rather than the within-school students  $n$ .

Note: As previously stated the equations noted above are to act as a guide for the evaluation. Part of the cost of the evaluation bid will be to work with the implementation team to establish the equation models that should be used in the final evaluation. It is likely that a three-level model may be used for analysis: pre to post test changes for students, in classes, in schools once discussions have finalised.

### **Evaluation team**

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

Dr Cary Roseth, Ass Professor of Educational Psychology, Michigan State University & 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 will design the equation for modelling data prior to the trial and will undertake HLM analysis of the final data set blind to condition.

Dr Liam O'Hare, Senior Research Fellow in Centre for Effective Education & Director Improving Children's Lives Project Queen's University Belfast. Liam has substantial experience as a principal investigator on a range of RCTs and Cluster RCTs with associated process evaluations. He has completed five 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 will help preparing the final report.

Other staff from Centre for Effective Education will be involved in the project, including a new appointment of a full-time research assistant. This person will have a professional qualification (e.g. preferably speech and language therapy, or as a second choice, a teaching qualification or health care with careful training on how to administer the ACE test) and evidence of postgraduate study at Masters level. They will administer and code ACE tests for 360 students pre and post test (160 days), collate data for NGRT for 1,860 students at pre-test (10 days) and administer (30 days) and collate (10 days) data on NGRT for 1,860 students at post-test. In addition they will interrogate the NPDB to identify and collate KS1, KS2, SPAG tests and phonics assessment data for 1,860 students (20 days). They will merge this data onto a single dataset and undertake data cleansing in preparation for data analysis (10 days). They will undertake interviews with 20 teachers, 20 members of senior management team and 2/3 local authority managers (40 days). They will code and analyse interviews (20 days). They will devise (under supervision of Thurston & O'Hare) (5 days) and issue a postal questionnaire to 128 teachers (5 days). They will analyse data from the questionnaire and prepare a report of findings (40 days). Thurston will manage the day-to-day work of the research assistant with help from O'Hare. Other research fellows are also available to support the team at times of enhanced workloads. They will be involved in writing the final report (20 days). In addition

they will be asked to comply with University health and safety requirements in terms of staff induction and safety at work and there will also be times that they are required to travel to schools and venues (not shown above).

Administration and support staff will be involved in supporting the evaluation. They will be responsible for physically distributing measures, collating data and data input into data sets. In addition they will book accommodation and travel for staff and process finances and claims for reimbursement from EEF. It is estimated that this will be a significant burden to the project and an administrator represents a very cost-effective means of reducing the higher costs of the RA and academic staff involved in the project.

### Risk analysis and counter measures

A risk analysis of School of Education and CEE activity has been undertaken. This is presented below by means of establishing the potential risks to the funder and the controls and contingency measures that are in place to minimise these risks. One of the major benefits of EEF funding this proposed evaluation are the extensive and strong controls and contingency measures that The Queen’s University of Belfast will be able to provide. This adds security to the funding body and peace of mind that the proposal will be delivered on specification and on-time [† = I is Impact L is Likelihood: 5=high to 1=low]

No.	Risks	Current Controls	Net Risk*		Contingency Plans
			I <sup>†</sup>	L <sup>†</sup>	
1	Failure assessments to collect the required attainment data (particularly KS1 7 KS2 assessments which have alpha values below 0.7)	Standardised measures from GL have been selected. These have been extensively developed and piloted and have alpha values above 0.7. They should be able to detect effect if there is any.	5	1	The text version of the GL Group Reading Assessment and ACE would be used to assess the students. Similar measures will be used at post-test.
2	Failure of schools to administer assessments at the required time.	Project administrator will notify schools of the correct time to administer the pre-intervention assessments, monitor the return of data and remind schools who haven’t returned data to do so. Project staff will offer telephone and on-line support to schools who experience difficulties with administering assessments.	5	2	In exceptional circumstances: Travel is costed into the bid to collect ACE assessment data. The RA could administer the GL NGRT Assessment to a non-tested school as a final fail-safe.
3	Failure to recruit RA to administer test on time.	KS1 data will be used for random assignment. Unemployment is high in Northern Ireland and many graduates fit the job profile who cannot find work at present.	3	1	In exceptional circumstances, CEE staff on permanent contracts will be able to visit schools and administer tests.
4	RA administers assessment incorrectly	Training will be given to the RA. Intra-rater reliability of the RA will be tested using video to ensure fidelity.	2	1	If an administration error is discovered through intra-rater testing, children will be re-

					assessed
5	Fire or other damage to electronically-stored data in CEE	Daily back-up of servers is carried out. Full back-up for off-site storage carried out every week Data is double backed up from key CEE personnel on a weekly basis also.	2	1	Files would be restored and any lost electronic data would be retrieved directly from schools and/or email systems. Files can be restored from one of the two back-ups.
6	Temporary loss of key project staff due to illness	CEE employs 17 full-time and dozens of part-time 'field worker' staff including administrators, technical staff who are familiar with supporting schools to administer assessments. There are two full-time staff with dedicated time to the project. In the unlikely event of illness either could fulfil the role of the other. The School of Education at QUB employs a similar number of staff and has very experienced researchers who could provide temporary cover for staff.	1	2	Other members of the CEE team who are not working on this project would be recruited along with others in their respective departments with the necessary skills and experience
7	Permanent loss of key project staff	See 7. Whilst a temporary solution would be put in place, a recruitment process would follow.	1	2	Other members of the CEE team who are not working on this project would be recruited along with others in their respective departments with the necessary skills and experience
8	Failure to obtain informed consent from schools	The bid will receive ethical permission from the School of Education Ethics Committee, which will scrutinise practices. The nature and purpose of data collection will be explained schools in advance of their recruitment to the project. Clear participant information sheets will be provided early on in the project to all potential participants. The data collection methods have been utilised in a number of previous projects without the materialisation of this potential risk as an actual problem.	5	1	Meetings with schools would be held to reassure them of the nature and purpose of data collection.
9	Research assistant does not have	QUB best practice on recruitment will be followed.	4	1	All candidates will be thoroughly screened



	appropriate skills set to carry out work/behaves inappropriately in school.	References will be thoroughly checked. Criminal record background of the applicant will be checked with advanced disclosure.			before offers of employment are made. Training will be given on administration of tests and measures.
10	Control schools decide they no longer want to be part of the evaluation after randomisation.	Clear information will be given to control schools. There is an incentive of post-treatment availability of resources that offers an incentive to keep participating.	4	2	Other incentives could be offered after discussion with EEF. A waiting list of additional schools could be used to populate the sample.
11	Differential of student attrition over conditions.	The sample is significantly robust enough to deal with such attrition.	3	2	Thought may be given to how to deal with missing data in the final model.
12	Lack of study power for an Effect Size of 0.25.	Whilst previous studies have shown the sample size <i>should</i> be sufficient, Effect Sizes varied widely dependent upon context.	3	2	If Effect Sizes are below 0.25 then the assumption of the positive impact of language-based interventions on reading may explored on an evidenced base. No result would be an interesting finding in itself and help to inform policy.

## Ethics

All research will be conducted according to QUB School of Education Ethical Guidelines. Ethical consent will be obtained from the Ethics Committee before data collection is conducted. Informed consent will be obtained from participants. Once coded and entered onto a database, data will be made anonymous and held securely on a password-protected computer.

## References

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.

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

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.