

Evaluation Summary	
Age range	Year 8 (12-13 years)
Number of pupils	Pilot: 900; main: 8700
Number of schools	Pilot: 12; main: 58
Design	RCT
Primary Outcome	Progress Test in Science (GL Assessments)

BACKGROUND

Significance

Studies in the field of neuroscience have found increased engagement associated with uncertain reward, and that levels of dopamine response to reward can predict memory for facts. Dopamine levels rise between an uncertain reward being anticipated and resolution in terms of whether it arrives, which explains the attraction of games of chance. There is some educational evidence that uncertain rewards lead to improved outcomes (Howard-Jones & Demetriou, 2009; Ozcelik, Cagiltay & Ozcelik, 2013), however few studies have been conducted in a real-life environment so far. The approach is now considered ready for testing to determine its application in classroom settings.

Intervention

This project will test the impact of a game-based approach to whole-class teaching, developed by researchers at the University of Bristol and Manchester Metropolitan University (MMU), which uses uncertain rewards for correct answers. Questions, posed using an online platform, will be integrated with class teaching in Year 8 science lessons, and students will work in teams to answer these. The study will explore three approaches to learning content: game-based (questions with uncertain rewards, where points are awarded for correct answers but teams can choose whether to keep their points or to risk doubling or losing them based on the chance spin of a wheel); test-based (questions with fixed rewards, i.e. a pre-determined number of points for being correct) and conventional teaching (teacher's usual practice). Although the points are not linked to any material reward, the project team hypothesises that the state of heightened excitement over whether or not they will gain or lose points in the "uncertain rewards" condition will lead to neuro-physiological changes that will increase their receptivity to learning.

The science curriculum is not taught in a consistent order across schools, and the position is further complicated because some schools complete KS3 in two years and others take three years. Bearing in mind these complexities, the project team aims to adapt the majority of KS3 science topics to the game-based and test-based models. Those not included (estimated at 20% of the curriculum) will be the areas least likely to be covered in Year 8.

There will be an initial phase of development work with schools to understand how the approach can be implemented most effectively in classrooms, and also to pilot evaluation methods and instruments. On the successful conclusion of this phase, a full-scale randomised controlled trial (RCT) will begin in September 2016.

ASSUMPTIONS

The IEE will work with the project team to draw up a theory of change. This model may further evolve in response to implementation and evaluation experiences, which may have implications for the research.

To allow cost and statistical calculations, we have assumed in this protocol that there are 25 students per class. In the main trial, the assumption is that each school will have six classes and approximately 150 students in Year 8. We acknowledge that reality will be more complicated, with some schools not participating with full cohorts. We have also assumed that the intervention will not interfere with topic content.

Although the project team will lead the recruitment, the evaluators will be closely involved. For instance, a member of the evaluation team will attend recruitment conferences to explain the RCT philosophy and process to potential participants. They will emphasise aspects such as the random nature of allocation, the importance of adherence to treatment and remaining in the evaluation until post-testing is complete. A shared document will be set up (using Google or similar) to track progress with recruitment.

A memorandum of understanding (MOU) will be signed by each school. This will outline the responsibilities of each main party (project team, evaluator and school). Before schools can be accepted onto the trial, they must have supplied a list of the names and Unique Pupil Numbers of all their Year 8 students and signed and returned the MOU to the IEE. This will facilitate pre-testing and applications for data (e.g. KS2 results) to the National Pupil Database (NPD). (NB this will not be so time critical in the main trial if it is decided to use a pre-test/post-test approach rather than using KS2 as a baseline). Schools will also have to administer an opt-out consent process by delivering letters to parents and registering any responses during a one-week period prior to randomisation.

RESEARCH PLAN: DEVELOPMENT AND PILOT PHASE

During this stage, the IEE will work closely with Bristol/MMU to share experiences and discuss possible adjustments to delivery, design or measures.

Research questions

The development phase will include a pilot RCT and will help determine the detail of the full RCT. It will allow us to test out the approach for recruitment, trial design, suitability of outcome measures and logistical issues alongside the final programme refinements being made by the project team. In particular, we can explore:

- whether KS2 results provide an adequate pre-test (if trial is conducted in England only) by comparing with a science pre-test
- what secondary measures are most appropriate
- how best to maximise response rate and optimise response quality with the measures employed

Design

At least three classes within twelve schools (approximately 36 Year 8 classes and 900 students) will be recruited to the pilot RCT. This is sufficient to allow trialling of measures and assessment of other aspects of the main trial. A minimum of three teachers per school should be involved.

Randomisation will take place at class level within each school, to divide the Year group into one game-based, one test-based and one control class (see Figure 1). Note: it is important that teachers are assigned to classes *before* randomisation to prevent potential bias if

teachers can choose their class knowing which treatment it is due to receive. Classes will be assigned to condition using either simple randomisation techniques if science is taught to mixed ability groups, or stratifying by ability if not (see p6 for more detail). The pilot will run for two terms in total, but the evaluation will only relate to the first term i.e. the spring term (see timeline on p9).

Figure 1: Pilot RCT

Pilot Study: 12 Schools, 3 classes per school (36 in total), 75 students per school (900 in total)

Treatment	School X		
	Control	Intervention A (games-based)	Intervention B (test-based)
	Class 1 25 Students	Class 2 25 Students	Class 3 25 Students

Participants

The schools will be English and Welsh state secondaries located near Bristol or Manchester for convenience for the development team. They will be chosen to represent the types of school we expect to recruit for the main RCT (e.g. higher than average FSM; mix of urban and rural; range of socio-economic and ethnic backgrounds).

Outcome Measures

The Progress Test in Science (published by GL Assessment) will be the primary measure. There is a version for administration at the start of KS3 (GL advise this should be used in the first two months of Year 7 as it is based on the primary science curriculum) and a version for the end of KS3 which exists in two forms: one schools that complete KS3 by the end of Year 8, and one for those that do not do so until the end of Year 9. We will use the Year 8 test.

Possible secondary measures will be trialled by the IEE and the project team for inclusion at the end of the main trial. These will include self-reported data on enjoyment, engagement and other attitudes towards science; attitudes to risk etc. Appropriate measures will be decided in consultation with the project team after a review of the literature, but are likely to include an adaptation of the attitudes to science measure used for the EEF evaluation of Thinking, Doing, Talking Science; attitudes, literacy and experience regarding gaming/digital games; and attitudes towards competition.

Half the students per treatment group in each school will complete the science test, and the other half will complete the secondary measures.

Analysis plan

Analysis will be conducted to calculate the correlation between the GL pre-test and the GL post-test, and between the KS2 SATs (for English, Maths or a combination) and the GL post-test. These will be compared and a decision made as to whether the GL pre-test or some form of KS2 data would provide the most appropriate pre-test, bearing in mind correlation levels and financial implications.

The secondary outcome measures will be examined for differences between the three groups, to establish which instruments should be used in the full trial. This decision will be based on a combination of factors, including apparent sensitivity to the intervention and ease of administration.

We understand that tracking data on students' use of the online platform will be available and we will include this in our statistical analyses to investigate whether the extent of use of the platform has any correlation with outcomes.

Recent experience has shown that many secondary schools struggle with online testing, so paper copy versions of the tests and questionnaires will be used.

Process evaluation methods

The development phase will allow the project team and evaluators to carry out joint observations and to construct an implementation rating scale for use in the main trial. The evaluators will have sight of the online training package. They will visit six of the schools, to observe lessons and to interview teaching staff about their experiences during the pilot, including the viability and practicability of the interventions.

At this stage, control classes will also be observed to determine whether such observations would be a useful element of the main trial.

Teachers and students will also be asked to complete an online survey. Questions will be piloted for use in the main RCT, and the survey will be developed in consultation with the project team. The teacher survey may include:

- general teaching behaviour and attitudes to teaching science (including controls)
- lesson preparation required (including controls)
- opinion of training/support provided
- how and where questions used in lesson
- usefulness of questions in assessing student performance
- perceived response of students to intervention
- overall opinion of intervention: strengths, weaknesses and suggestions for improvement
- likelihood of continuing with the approach/adopting it with other year groups
- awareness of other approaches being used in school in Year 8 science (to explore overlap/diffusion) (including controls)

Students taking part in each of the three arms of the study will be asked a short series of questions, including:

- enjoyment of/interest in science lessons (a proxy for engagement that has been shown to influence subsequent uptake of science)
- use of teaching approaches (including quiz questions in intervention)
- engagement with group work
- awareness of other approaches being used in school in Year 8 science (to explore diffusion between treatment groups)
- (intervention arms only) overall opinion of intervention: strengths, weaknesses and suggestions for improvement

After data collection has been completed, at least one member of staff from each of the schools will be interviewed by telephone or email to explore any issues that may have arisen during the testing process and to gather any suggestions for improvements in the full trial.

Formative findings from the pilot will be fed back on an ongoing basis to the project team as appropriate, to input into programme developments and design refinements for the main trial.

RESEARCH PLAN: MAIN RCT

Research questions

The full trial is designed to establish:

- what is the impact of the game-based teaching versus conventional teaching on academic achievement in science?
- what is the impact of the test-based teaching versus conventional teaching on academic achievement in science?
- what is the impact of the game-based teaching versus test-based teaching on academic achievement in science?
- what is the impact of the different approaches on secondary outcomes (to be finalised during the pilot, e.g. attitudes towards and engagement with science, attitudes to risk)?
- how are the two approaches enacted and received in the classroom, and how does this compare with “business as usual”?

Design

The details of the design depend on the results of the pilot. As in the pilot, teachers should be assigned to classes *before* randomisation to prevent potential bias resulting from prior knowledge of which classes will receive which treatment. Randomisation will be conducted at class level and all three conditions will run within each school.

The advantages of this approach are that individual schools will teach science content at roughly the same time so this should not be a variable in the test outcome; all schools will be motivated by being part of the intervention as none will be control-only; and a smaller sample size is needed. It also guarantees a reasonably even distribution across treatment arms between those students covering KS3 in two years and those completing it over three years, and increases the likelihood that the order of coverage of topics will be similar between treatments. Both these factors could have an effect on the performance on a science outcome measure that would be unrelated to the interventions.

However, the design carries some risks. There may be diffusion between the treatments; teachers may be confused if they have to teach using more than one approach; and the design has to take account of streaming and setting in schools. The disadvantages are all manageable with careful design and briefing to schools and teachers.

Power analysis calculation

A sample of 58 schools will be needed to achieve an MDES for a comparison between any of the arms of 0.1 standard deviations at 80% power.

Students per school per class: 25 (i.e. 50 per treatment per school)

Pre-post correlation (squared): +0.49

Rho (variation between classes within school): 0.05

Intraclass correlation: 0.125

Sigma² (effect size variability): 0.01

Criterion for statistical significance: $p < .05$

Power: 0.80

MDES = 0.1

At 0.1, the MDES is much lower than would be typical, but this takes into account that there will be three comparisons: Control (C) vs Intervention A; C vs Intervention B; and A vs B. It is hypothesised that the difference between interventions A and B could be much smaller than between each intervention and the control, which is why a lower MDES has been put into the power calculation.

This calculation assumes the entire Year 8 cohort will be involved in the study at all participating schools. The project team is concerned that some schools will be unwilling to be involved on this basis. If necessary to achieve the recruitment target, a school will be allowed to involve a part-cohort, but a minimum of three classes and three teachers must be involved. If some schools have fewer than two classes and 50 students per treatment arm (either because of their small size or by choice), the power will be reduced accordingly.

Randomisation

The approach to randomisation will vary depending on whether schools stream or set by ability in science lessons. For schools that do stream or set, the sample will be stratified by ability to ensure that across those schools, an even number of high, medium and low ability classes are assigned to each treatment. If schools teach science in genuine mixed ability groupings, simple randomisation techniques will be used.

Participants

State secondary schools in more than one geographical region will be recruited. They should have a higher than average proportion of students on free school meals (FSM), be drawn from a range of contexts and be at least 3-form and ideally 6-form entry.

Outcome Measures

The Progress Test in Science (published by GL Assessment) will be the primary measure. The secondary measures will be selected from those trialled in the pilot RCT, and are expected to include an attitudinal measure focused on science and an assessment of attitudes to risk.

Analysis plan

Multi-level modelling will be used with a covariate confirmed by the pilot trial (Progress Test in Science or KS2 results in some form). Sub-group analysis will be conducted on FSM students. Gender (and science ability based on pre-test, if administered) will be used as factors in the model, including tests for interaction. Additional factors, for instance online trace data as proxy for student engagement and student attitudes to risk, will be used if warranted by data from the pilot. We will analyse the pre-test data to check for balance at baseline.

The effect size will be calculated using Hedges' g . We will use the adjusted difference in the means between the arms divided by the unadjusted total standard deviation in the outcome at the end of the trial.

Process evaluation methods

The process evaluation will be based on practices and instruments trialled and refined in the development phase. If online training only is provided to teaching staff, as currently anticipated, the evaluation team will have access to the materials. If there is an element of face-to-face training, at least one session will be attended by the evaluation team. It is recommended that 12 of the 58 schools are visited and one lesson from each condition observed in each. They would be visited twice: once earlier and once later in the intervention period to monitor any changes in perceptions. Interviews will be held with science staff about their experiences with the treatments, and with senior management for a wider perspective.

Teachers and students will be asked to complete an online survey, which will be designed in line with learning from the development phase.

PERSONNEL

Pam Hanley, PhD. Principal Investigator. Pam Hanley is a Research Fellow at the IEE and has a solid track record in science education research and the use of both qualitative and quantitative methods. She has worked on a wide variety of evaluations including randomised controlled trials in science, numeracy and literacy. Science education research is one of her specialisms, and she has conducted RCTs focused on science outcomes at both primary and secondary level.

Bette Chambers, PhD. Project advisor. Bette Chambers is Director of the Institute for Effective Education at the University of York. Professor Chambers develops and evaluates effective practices in education and promotes the use of evidence-based practices. She has authored or co-authored numerous articles, books, and practical guides for teachers, including *Let's Cooperate* and *Two Million Children*.

Louise Elliott. Data Manager. Louise Elliott is the Data Manager at the IEE, where she manages all database organisation, data entry, cleaning and descriptive statistical analyses conducted in the research work. Louise is very experienced at implementing, administering and collecting data from school tests, including GL online assessments.

Project Assistant [TBC]. A Project Assistant will work on this evaluation. Supervised by the Principal Investigator, he or she will work on all aspects of the project, assisting the team with communications with schools, the organising of test papers, design of the teacher/student interview schedules and lesson observations, data collection, and the writing of reports.

Data protection statement: data will be handled in accordance with the Data Protection Act (1998). All outputs will be anonymised so that no schools will be identifiable in the report or dissemination of results. Statistical databases will hold non-identifiable data. Twenty per cent of data will be double entered to assess reliability where appropriate. Confidentiality will be maintained and no one outside the evaluation team will have access to the database. The trial database will be securely held and maintained on the University's research data protection server, or by GL Assessments. Ethical approval for this study will be sought through the Ethics Committee of the Department of Education, University of York.

RISKS

Risk	Preventative measures
Insufficient schools recruited	Use project team contacts, potential intermediaries (e.g. academy chains, local authorities), IEE and EEF outlets (e.g. website, Twitter, email lists).
Attrition (from programme)	Provide adequate explanation of programme in advance; sound training and support; foster good relations with individual teachers and schools.
Attrition (from evaluation)	Explain RCT design and expectations thoroughly at recruitment; minimise burden on schools, teachers and students; foster good links with schools.
Diffusion/spillover	Need to stress importance of not exposing students to any other than their allocated condition. Teachers to be clear what constitutes each treatment. If they are randomised to teach more than one treatment, they need to be carefully instructed then closely monitored and supported to ensure there is no overlap.
Low implementation fidelity	Project team to monitor trace data from platform provider and take action if necessary with individual teachers.
Technology problems	Schools to have rapid access to technical assistance from project team or platform provider.
Teacher turnover	Provide immediate training and support for new staff.
Changes in evaluation team	Procedures in place within IEE to enable other staff to take over quickly and efficiently if necessary.
Slippage in timing (recruitment)	Recruitment over a sufficient time period with interim assessments of progress.
Slippage in timing (evaluation)	Project assistant dedicated to arranging testing and chasing schools.

REFERENCES

- Howard-Jones, P. A. & Demetriou, S. (2009). Uncertainty and engagement with learning games. *Instructional Science* 37, 519-536.
- Ozcelik, E., Cagiltay, N. E. & Ozcelik, N. S. The effect of uncertainty on learning in game-like environments. *Computers & Education* 67, 12-20, doi:<http://dx.doi.org/10.1016/j.compedu.2013.02.009> (2013).

TIMELINE

PILOT RCT	Spring 2015	Summer 2015	Autumn 2015	Spring 2016	Summer 2016	Autumn 2016	Spring 2017	Summer 2017	Autumn/ Spring 2017/18
Work with project team to recruit schools for pilot									
Obtain UPNs and, via NPD, KS2 results									
Science pre-test									
Randomly assign classes to three arms of study									
Intervention									
School visits (x6): lesson observations & interviews with school staff									
Online teacher and student survey									
Science post-test and other measures									
Initial analysis of survey and post-data									
Phone interviews with participating school staff (x6)									
Report and recommendations based on pilot RCT									

MAIN TRIAL	Spring 2015	Summer 2015	Autumn 2015	Spring 2016	Summer 2016	Autumn 2016	Spring 2017	Summer 2017	Autumn/ Spring 2017/18
Work with project team to recruit schools for main trial									
Obtain UPNs (and, via NPD, KS2 results)									
Randomly assign classes to three arms of study									
Observe training (if relevant)									
Science pre-test									
Intervention									
School visits: lesson observations & interviews with school staff									
Online teacher and student survey									
Collect and analyse post-intervention data (GL Science and secondary measures)									
Write report									