

Evaluation Summary	
Age range	Secondary (Year 9 – 11)
Number of pupils	c. 7,600
Number of schools	40
Design	Randomised controlled trial with randomisation at the school level
Primary Outcome	Maths and English GCSEs

Evaluation Protocol

Independent evaluation of the ‘Inclusive’ project

1.1 Introduction

We are delighted to have the opportunity to lead this evaluation, which fits perfectly with our ongoing programme of research on school-based social and emotional learning (SEL) interventions. We led the national evaluations of the secondary SEAL programme (Humphrey, Lendrum, & Wigelsworth, 2010), the primary SEAL small group work element (Humphrey et al., 2008), and the Achievement for All initiative (which included strands on improving behaviour, positive relationships and bullying - Humphrey & Squires, 2011). We are currently leading a major cluster randomised trial of the PATHS curriculum that is being part-funded by the EEF. Additionally, Humphrey was a Co-I on the national evaluation of the Targeted Mental Health in Schools evaluation (led by UCL) (Wolpert et al, 2011), which included a randomised trial within the overall design. More broadly, our team has considerable expertise in the implementation of school-based interventions (Lendrum, Humphrey, & Wigelsworth, 2013; Lendrum & Humphrey, 2012) and assessment of social and emotional skills (Humphrey et al., 2011; Wigelsworth, Humphrey, Kalambouka, & Lendrum, 2010). Finally, a number of the projects we have led or collaborated on have involved the collection of naturally occurring academic attainment data (e.g. GCSE scores) (the PATHS and AfA projects as part of the original design; secondary SEAL and TaMHS as part of follow-up work).

1.2 The evaluation team

Our evaluation team brings together both the methodological and subject expertise required to produce a rigorous evaluation of the Inclusive project. The team members are:

Dr. Michael Wigelsworth (PI) will direct the study and ensure that it is completed to time and budget. Michael specialises in measuring the effects of school-based interventions. Michael is currently leading the outcome component of our on-going randomised controlled trial of the Promoting Alternative Thinking Strategies (PATHS) curriculum. He brings particular expertise in quantitative methods and statistical modelling.

<http://www.manchester.ac.uk/research/michael.wigelsworth/>

Prof. Neil Humphrey (Co-I) has considerable experience in leading large-scale evaluation trials in educational settings (see above) and is an expert on social and emotional learning.

He has researched and written extensively in this area, including a recent major text described as “a landmark for the field” (Professor Roger Weissberg, University of Illinois at Chicago) (Humphrey, 2013).

<http://www.manchester.ac.uk/research/neil.humphrey/>

Dr. Ann Lendrum (Co-I) is an expert in the evaluation of implementation processes in school settings. Ann is currently directing the implementation strand of the aforementioned PATHS trial, and led the process components of the national evaluations of the Achievement for All and secondary SEAL programmes. She is a founding member of the UK Implementation Network.

<http://www.manchester.ac.uk/research/ann.lendrum/>

The team will also include experienced research staff to supervise the data collection. Among our current roster are Drs. Alexandra Barlow and Afroditi Kalambouka, both of whom have experience of working on large-scale educational trials, administration of academic assessments, and managing junior researchers.

1.3 Our understanding of the Inclusive programme, its rationale and context

From the information provided to us, Inclusive can be described as a multi-component universal SEL intervention. Using the classification system adopted in recent major reviews in this field (e.g. Blank et al., 2010), it combines *curricular* and *environmental* components. From the perspective of Humphrey’s (2013) SEL taxonomy, it may be described as a hybrid programme in terms of its prescriptiveness, offering both ‘manualised’ content of core components and flexible, needs-led delivery of non-core components. Interventions such as Inclusive can be seen as part of a growing attempt to make schools central to efforts to improve the mental health and wellbeing of students in the United Kingdom. Although it appears to have stabilized in the new millenium, the decades prior to this saw marked increases in ‘externalising’ behaviour problems (e.g. aggression, conduct problems) among children and young people (Maughan, Iervolino, & Collishaw, 2005), with higher rates of disorder among adolescents compared to children (Green, McGinnity, Meltzer, Ford, & Goodman, 2005). Additionally, a recent, large-scale survey indicated that 1 in 10 students report being bullied every day (Chamberlain, Nalia, Golden, Walker, & Benton, 2010).

By virtue of their wide reach and central role in the lives of children and young people, schools are ideally placed to be at the heart of efforts to prevent such problems (Greenberg, 2010). One approach is the implementation of universal SEL programmes such as Inclusive. Two recent meta-analyses have provided strong evidence that high quality, well implemented interventions of this nature can yield positive effects on both psychosocial and academic outcomes (Durlak, Weissberg, Dymnicki, Taylor, & Schellinger, 2011; Sklad, Diekstra, De Ritter, Ben, & Gravesteyn, 2012). However, in the English context school-based prevention efforts have met with limited success thus far, particularly in secondary education. Our team’s national evaluation of the secondary SEAL programme suggested that it had failed to impact upon student outcomes (Humphrey, Lendrum, & Wigelsworth, 2010). Since the Coalition government’s subsequent withdrawal of their endorsement of the SEAL programme (BBC, 2011) and other actions that have signified a shift away from New Labour’s emphasis on ‘emotional education’ (e.g. the removal of ‘personal development and wellbeing’ from the school inspection framework; Ofsted, 2012), there has been something of a void in this area. This is particularly true in the secondary sector, where a more rationalist approach to schooling

can mean that SEL interventions are a ‘harder sell’ and often meet greater attitudinal and logistic challenges to implementation than in primary education (Lendrum et al., 2013). The impact of new, evidence-informed initiatives (such as How To Thrive, funded through the EEF) that have begun to appear are not yet known.

Inclusive – which has already undergone a pilot trial (www.inclusiveschools.org.uk)¹ - is to be launched into the above context and evaluated rigorously through a major cluster randomised trial funded by the National Institute for Health Research. Our understanding of the EEF’s involvement is that it is very similar to that of our own PATHS trial – ‘bolt-on’ funding of intervention costs and academic measurement that can be integrated into the existing trial protocol. The role of the independent evaluation, then, is to determine the impact of Inclusive on the academic attainment of pupils in participating schools.

2.1 Our evaluation strategy

Our independent evaluation of Inclusive benefits from the very clearly defined trial protocol that is already in place for the main trial. In short, 40 secondary schools in England will be recruited and randomly assigned to implement the Inclusive intervention over 3 years or continue usual practice during the equivalent period of time. Schools in the intervention arm will receive technical support and assistance from external facilitators for the first 24 months. The target cohort comprises pupils in Year 8 (e.g. aged 12/13) at the outset of the intervention. Primary outcome measures of aggression and bullying will be taken at baseline and 36 month follow-up, with secondary outcome measures at school and student levels taken at 24 and 36 months. A process evaluation examining different aspects of implementation (e.g. fidelity/adherence) will take place throughout the main trial period.

The current trial is more than adequately powered to detect effects on the primary outcome measures for which it was originally designed. However, measures of academic attainment present a quandary. First, the ICC for attainment in secondary schools (approximately 0.27²) is much larger than for aggression or bullying (which has been specified at 0.04). Second, the expected effect size (ES) for attainment is presumably much smaller than for aggression or bullying (which has been specified at approximately 0.23), as the former is presumably an indirect, distal outcome of the main intervention processes and effects. So we might reasonably expect an ES for attainment as low as 0.1. The compound effect of these factors means that the trial is underpowered to detect effects on attainment by some considerable degree, as follows:

Assuming $N=190$ per cluster, $ICC=0.27$, $ES=0.1$, $Power=0.8$ and $Alpha=0.05$, approximately 860 clusters/schools would be required. Even if the ES were aligned with that expected for the primary outcome measure (e.g. $ES=0.23$), approximately 170 clusters/schools would be needed.

Increasing the number of clusters is clearly the most efficacious strategy, but even the additional 5 per arm indicated in the correspondence relating to this project would not bring us anywhere near the requirements outlined above. Increasing the N per cluster by including additional year group cohorts from existing schools, as has also been indicated, would reduce the number of clusters required, but again not by any meaningful degree. For example, one

¹ An obvious starting point for us would be to examine the findings of this trial, which are due to be published in February 2014. Although this trial did not include assessment of academic progress, other measures of school outcomes (e.g. attendance and exclusion) may give a useful indication of whether impacts on proximal variables such as aggression and bullying have triggered distal effects.

² This figure was calculated using GCSE scores for English and Maths in the National Pupil Database (NPD).

additional year group cohort (making N=380 per cluster) would mean 162 clusters would still be required to detect an ES of 0.23 on attainment. An additional issue is that increasing the number of clusters and/or N per cluster has knock-on effects for the main project team vis-à-vis their intervention delivery capacity (e.g. do they have access to the staff and budget to facilitate the intervention in a larger number of schools and/or to larger numbers of pupils³?)

In light of the above, we do not see any immediate solutions to this quandary. Currently, the existing design is powered thus:

Assuming N=190 per cluster, 40 clusters, ICC=0.27, Power=0.8 and Alpha=0.05, an ES of 0.48 or larger would be detectable.

This is one potentially useful benchmark as it aligns closely with the average ES of 0.46 for academic attainment found by Sklad et al's (2012) aforementioned meta-analysis of the impact of universal SEL interventions. However, experience from previous projects suggests that the ICC for attainment in secondary schools in evaluation trials may be lower than that found nationally. This is because the initial recruitment pool is more homogenous than the national picture – for example, all schools are drawn by a common interest in the intervention in question. In the aforementioned secondary SEAL evaluation (Humphrey, Lendrum & Wigelsworth, 2010), the ICC for the attainment was 0.21 for the 41 participating schools. Given this, a more accurate assessment of the power status of the Inclusive trial is as follows:

Assuming N=190 per cluster, 40 clusters, ICC=0.21, Power=0.8 and Alpha=0.05, an ES of 0.41 or larger would be detectable.

This is another useful benchmark as it corresponds directly to Hattie's (2009) 'hinge point', at which, "the effects of innovation enhance achievement in such a way that we can notice real-world differences" (p.17). Thus, rather than seeking to *manipulate and extend* the existing design in pursuit of the ability to detect a small effect (in the context of a larger ICC), we will *retain* the existing design. This approach will allow us to ask two critical questions: (i), does Inclusive produce effects on attainment that are comparable with those of existing SEL programmes (following Sklad et al, 2012)? and (ii) does Inclusive produce effects on attainment that are meaningful (following Hattie, 2009)?

In terms of what data is collected, administration of independent standardized assessments is the most rigorous and objective approach. Products such as Yellis, produced by CEM in Durham, would yield the data required and offer the advantage of providing detailed and valuable individual and aggregated feedback to schools. However, this comes at considerable cost, both in financial terms and data burden to schools (with the subsequent risk of increased noncompliance and/or attrition this brings).

Instead, we intend to use naturally occurring academic data as this is the most viable and efficient option, in addition to adding considerable value in terms of the questions it would allow us to answer (see below). All participating schools will produce data through Key Stages (KS) 3 and 4 that could be retrieved and used as outcome data in our independent evaluation with no or minimal burden to schools. First, when the target cohort reach the end of Year 9 (2016), they will be subject to teacher assessment judgements that are submitted to (and could easily be retrieved from) the National Pupil Database (NPD). Second, when they

³ We recognize that much of Inclusive is delivered at the 'whole school' level; however, any increase in the number of pupils in participating schools will always have implications for delivery capacity because of the SEL curriculum component.

reach the end of Year 10 (2017), further teacher assessment judgements are typically generated to feed forward to GCSE examination entry decisions. Although the results are not submitted to the NPD, they could be retrieved directly from schools with minimal burden. Third, when the target cohort complete Year 11 (2018), they will sit their GCSE examinations. As with their KS3 data, this could be retrieved from the NPD. The timing of these assessments allows us to answer several important questions, as follows (end of KS2 attainment data operates as a baseline co-variate as appropriate):

Using Year 9/KS3 data

- Efficacy: What is the impact of Inclusive on the academic attainment of pupils given optimal conditions (e.g. following 2 years of external intervention facilitation)?

Using Year 10/GCSE 'prep' data

- Effectiveness: What is the impact of Inclusive on the academic attainment of pupils given typical conditions (e.g. following 1 year *without* external intervention facilitation)?

Using Year 11/GCSE data

- Follow-up: Is any impact of Inclusive on the academic attainment of pupils maintained at follow-up? *or*

- Sleeper: (assuming no impact at post-test) Is there a sleeper effect of Inclusive on the academic attainment of pupils?

3.1 Deliverables

In light of the discussion with the EEF and Inclusive project team, Manchester will be responsible for collecting the academic data, providing analysis and authoring the EEF report. As part of the NHIR main trial, IOE will be collecting process data, which will be shared with Manchester to augment the IIT analysis, specifically; a) The use of the annual researcher observation data and follow up surveys from staff and students (participation, reach and dose) as explanatory variables of academic outcomes and b) The use of interview data from case study schools to explore potential explanatory mechanisms of actions leading to outcome data. Both approaches will be considered in conjunction with the proposed analysis of temporal relations (see below) in attempt to establish an 'evidence chain' between theory, implementation, and outcome.

Once data sharing is formally made part of the ethical protocol for the study (e.g. participants give consent for their data to be shared with a third party) Manchester will include data gathered from the Inclusive Project team, as specified below. We will obviously work closely with the project team to ensure that the research meets the requirements of the various stakeholders (e.g. EEF, NIHR, participating schools).

It is agreed that Manchester will report on attainment as the primary outcome, specifically Maths & English GCSE results. We will conduct intention-to-treat analyses using hierarchical linear modeling to take into account the clustered and hierarchical nature of the study dataset. ES will be reported using Hedge's g (Cohen's d bias corrected) and accompanied by 95% confidence intervals as per EEF specifications.

Secondary outcomes will include additional analysis of Yr 9 and Yr 10 Maths and English (in order to answer the additional questions posed above). Subgroup analysis (utilising Inclusive data) will consider the influence of free school meal eligibility, gender, and victims and perpetrators of bullying (as identified using Inclusive project data). An additional consideration

is the differential effects for adolescents presenting in the clinical, subclinical and normal ranges for aggression at baseline. These kinds of additional analyses have been called for in the SEL literature recently (e.g. Durlak et al, 2011; Humphrey, 2013) and would greatly strengthen our understanding of the Inclusive intervention and the processes by which it may impact upon outcomes for young people. There will also be an adherence to protocol analysis looking at the moderating effect of implementation (e.g. fidelity and dosage, using IOE data). This will be vital in examining whether variations in findings may be attributable to a lack of implementation fidelity.

Manchester will also consider temporal relations between outcome variables that might help to empirically validate the intervention logic model. We note that the logic model included in the pilot trial protocol (http://www.nets.nihr.ac.uk/_data/assets/pdf_file/0017/53135/PRO-09-05-05.pdf) specifies student health rather than academic outcomes as the end point in the causal chain. Although a generic model could be applied from extant theory (see for example, CASEL, 2007) and empirical evidence (e.g. Durlak et al, 2011; Sklad et al, 2012) pertaining to the influence of SEL interventions on attainment, we feel that this project merits the explicit, apriori development (and subsequent empirical validation) of a reworked logic model for Inclusive that takes into consideration attainment as a distal outcome variable. This kind of work has been integral to our activities on other projects (e.g. Lendrum, 2010). Crucially, it will allow an assessment of the likelihood of ‘theory failure’ in the event of null results.

Academic publications arising from these analyses will be co-authored between research teams.

4.1 Timescale

Month	Inclusive Project Delivery	Management and administration	Fieldwork/data collection for independent evaluation	Analysis and writing up
2013 Nov/Dec		Meeting with EEF and project team		
2014 Jan/Feb Mar/Apr	- Recruitment of schools, fieldworkers and consultants - Instrument preparation		NPD extraction: cohort data	
May/June	Baseline (T1) outcome measures	Virtual (Skype) progress meeting with project team		Progress report to EEF
July/Aug	Randomisation			
Sep/Oct	Facilitated intervention begins	Start of efficacy period		Progress report to EEF
Nov/Dec				
2015 Jan/Feb		Virtual (Skype) progress meeting with project team		
Mar/Apr				Progress report to EEF
May/June				

July/Aug		Progress meeting with project team		
Sep/Oct				Progress report to EEF
Nov/Dec				
2016		Virtual (Skype) progress meeting with project team		
Jan/Feb				
Mar/Apr				Progress report to EEF
May/June	Interim (T2) outcome measures	End of KS3 Teacher assessment judgements		
July/Aug		Progress meeting with project team		
Sep/Oct	Non-facilitated intervention begins	Start of effectiveness period		Progress report to EEF
Nov/Dec				
2017		Virtual (Skype) progress meeting with project team	NPD extraction: KS3 data	
Jan/Feb				
Mar/Apr				Progress report to EEF
May/June	Final (T3) outcome measures	Year 10 Teacher assessment judgements	Extraction of teacher assessment judgements	
July/Aug		Progress meeting with project team		Analysis and writing of main report
Sep/Oct		Start of follow-up/sleeper period		
Nov/Dec				Main report to EEF 18/12/2017
2018		Virtual (Skype) progress meeting with project team		
Jan/Feb				
Mar/Apr				Progress report to EEF
May/June		Cohort sit GCSE examinations		
July/Aug				
Sep/Oct		Virtual (Skype) progress meeting with project team		Progress report to EEF
Nov/Dec				
2019			NPD extraction: GCSE data	Analysis and writing of follow-up report
Jan/Feb				Follow-up report to EEF 28/02/2019

References

5.1 Timescale

- BBC. (2011). Analysis - testing the emotions. Retrieved from <http://www.bbc.co.uk/programmes/b00z5bq4>
- Blank, L., Baxter, S., Goyder, L., Guillaume, L., Wilkinson, A., S., H., & Chilcott, J. (2010). Promoting wellbeing by changing behaviour: a systematic review and narrative synthesis of the effectiveness of whole secondary school behavioural interventions. *Mental Health Review Journal*, 15, 43–53.
- CASEL. (2007). *How Evidence-Based SEL Programs Work to Produce Greater Student Success in School and Life*. Chicago: CASEL.
- Chamberlain, T., Nalia, G., Golden, S., Walker, F., & Benton, T. (2010). *TellUs4 National Report*. London: DFE.
- Durlak, J., Weissberg, R. P., Dymnicki, A. B., Taylor, R. D., & Schellinger, K. B. (2011). The impact of enhancing students' social and emotional learning: a meta-analysis of school-based universal interventions. *Child Development*, 82, 405–32.
- Green, H., McGinnity, A., Meltzer, H., Ford, T., & Goodman, R. (2005). *Mental health of children and young people in Great Britain*. Newport: ONS.
- Greenberg, M. T. (2010). School-based prevention: current status and future challenges. *Effective Education*, 2, 27–52.
- Hattie, J. (2009). *Visible learning*. London: Routledge.
- Humphrey, N., Kalamouka, A., Bolton, J., Lendrum, A., Wigelsworth, M., Lennie, C., & Farrell, P. (2008). *Primary social and emotional aspects of learning: evaluation of small group work*. Nottingham: DCSF Publications.
- Humphrey, N., Lendrum, A., & Wigelsworth, M. (2010). *Secondary social and emotional aspects of learning (SEAL): national evaluation*. Nottingham: Department for Education.
- Humphrey, N., Kalamouka, A., Wigelsworth, M., Lendrum, A., Deighton, J., & Wolpert, M. (2011). Measures of Social and Emotional Skills for Children and Young People: A Systematic Review. *Educational and Psychological Measurement*, 71, 617–637.
- Humphrey, N. (2013). *Social and emotional learning: a critical appraisal*. London: Sage Publications.
- Humphrey, Neil, Lendrum, A., & Wigelsworth, M. (2010). *Social and emotional aspects of learning (SEAL) programme in secondary schools: national evaluation*. Nottingham: DFE.
- Humphrey, N., & Squires, G. (2011). *Achievement for All national evaluation: final report*. Nottingham: DFE.
- Lendrum, A., Humphrey, N., & Wigelsworth, M. (2013). Social and emotional aspects of learning (SEAL) for secondary schools: Implementation difficulties and their implications for school-based mental health promotion. *Child and Adolescent Mental Health*, 18, 158-164.
- Lendrum, Ann. (2010). *Implementing Social and Emotional Aspects of Learning (SEAL) in secondary schools in England: issues and implications*. University of Manchester: Unpublished PhD thesis.
- Lendrum, Ann, & Humphrey, N. (2012). The importance of studying the implementation of school-based interventions. *Oxford Review of Education*, 5, 635-652.
- Maughan, B., Iervolino, A. C., & Collishaw, S. (2005). Time trends in child and adolescent mental disorders. *Current Opinion in Psychiatry*, 18, 381–385.
- OFSTED (2012). *The framework for school inspection: Guidance and grade descriptors for inspecting schools in England under section 5 of the Education Act 2005, from January 2012*. Manchester: OFSTED.
- Sklad, M., Diekstra, R., De Ritter, M., Ben, J., & Gravesteyn, C. (2012). Effectiveness of school-based universal social, emotional, and behavioral programs: do they enhance students' development in the area of skills, behavior and adjustment? *Psychology in the Schools*, 49, 892–909.
- Wigelsworth, M., Humphrey, N., Kalamouka, A., & Lendrum, A. (2010). A review of key issues in the measurement of children's social and emotional skills. *Educational Psychology in Practice*, 26, 173–186.
- Wolpert, M. et al. (2011). *Me and my school: findings from the national evaluation of Targeted Mental Health in Schools*. Nottingham: DFE.