Trial Evaluation Protocol
Adventure Learning

Evaluator (institution): Sheffield Hallam University
Principal investigators: Professor Tim Jay

PROJECT TITLE	**Adventure Learning**
DEVELOPER (INSTITUTION)	The Outward Bound Trust and Commando Joe's
EVALUATOR (INSTITUTION)	Sheffield Hallam University
PRINCIPAL INVESTIGATOR(S)	Tim Jay
PROTOCOL AUTHOR(S)	Tim Jay, Sarah Reaney-Wood, Ben Willis, Sean Demack
TRIAL DESIGN	Three-arm clustered randomised controlled trial with random allocation at school level
PUPIL AGE RANGE AND KEY STAGE	Year 9 pupils (13-14 years old), KS3 at time of intervention
NUMBER OF SCHOOLS	99
NUMBER OF PUPILS	2376
PRIMARY OUTCOME	Self-regulation of learning (SRL-SLS)
SECONDARY OUTCOMES	Student behaviour (Strengths and Difficulties questionnaire, SDQ),Student engagement (SEI); GL Progress test in maths, Attainment 8

PROTOCOL VERSION HISTORY

VERSION	DATE	REASON FOR REVISION
1.0	15 February 2019	N/A

TABLE OF CONTENTS

Intervention description

Adventure learning ('the Programme') is a year 9 intervention the evaluation of which will research the elements that are most important in the development of non-cognitive and academic outcomes (for example, intense, week-long experience; challenging adventure; engagement with nature) and how the outcomes associated with these programmes (for example, increased self-efficacy, self-regulation and resilience; improved relationships in school; behaviour and attitudes in the classroom) may link to improved attainment.

The Programme will involve two established organisations in this field: The Outward Bound Trust (OBT) and Commando Joe's (CJ's). In the Outward Bound group, pupils will take part in challenging, adventurous activities such as canoeing, hiking and wild camping, as part of an intensive five-day residential course delivered in one of five locations across England, Scotland, and Wales. Instruction will be delivered by trained outdoor learning facilitators in collaboration with accompanying staff from the pupils' school. Learning strategies such as growth mindset theory, goal-setting, reflection and feedback will be used by instructors during the course to enhance learning. Pupils will take part in groups of 12 (two groups per school) with each group led by a qualified OBT instructor, and accompanied by a teacher from the pupils' school, for the 5-day programme. Groups of 12 are required for this intervention as OBT residential programmes are designed for group of this size.

Pupils in the Commando Joe's group will similarly combine challenging physical activity with the use of metacognitive skills and instructor-facilitated reflection sessions aiming to improve non-cognitive outcomes and attainment. Commando Joe's trained instructors are military veterans, and this programme will be delivered over five consecutive days on the school site. Pupils will take part in this intervention in a group of 24, led by a CJ's instructor and accompanied by two teachers from the pupils' school. This number has been chosen to match the number of pupils taking part in the OBT condition.

Both interventions last for 5 days. The OBT intervention is a residential experience at one of five Outward Bound centres in the UK. The Commando Joe's intervention will take place over 5 consecutive days at the pupils' school, making use of the school building and grounds. Both programmes broadly follow a standard model of delivery, but there is some tailoring to the particular needs of pupils based on pre-intervention meetings between delivery staff and teachers in schools.

Neither intervention demands that teachers carry out specific activities before or after the 5-day intervention. However, both require that 2 teachers[1] from each school accompany the 24 pupils, and both interventions have a resource pack for teachers to support further work with pupils post-intervention.

Schools in the control group will receive a financial incentive of £1500, paid in three instalments following key data collection points. The expectation is that after coving any costs, this money will be put towards providing activities to enrich educational experiences, with specific focus on the 24 pupils identified to take part in the trial.

Study rationale and background

Effectiveness of outdoor adventure learning

Meta-analyses have provided some evidence for the effectiveness of outdoor adventure learning on raising attainment (Cason & Gills, 1994; Hattie et al., 1997). These meta-analyses reveal average effect sizes in the region of 0.30, representing a moderate effect[2]. The evidence in this domain requires some updating, as available studies in this domain are rather dated. There is also a need for

[1] As agreed after further consideration at the IDEA workshop.
[2] an effect size of 0.30 falls within the 0.19 to 0.44 'moderate' range in the EEF DIY evaluation guide (https://v1.educationendowmentfoundation.org.uk/uploads/pdf/EEF_DIY_Evaluation_Guide_(2013).pdf)

research that can shed light on mechanisms of improvement in attainment. Hattie et al. (1997) report an unusual effect whereby attainment continues to improve (with an additional effect size of 0.17) after the end of the intervention and before follow-up analyses. This finding points to the importance of follow-up actions in schools, and the development over time of contributing factors to raised attainment. This in turn is supported by a relatively large body of research describing improvements in non-cognitive outcomes following outdoor adventure learning interventions.

Hattie et al. (1997) report that effect sizes vary widely across different interventions, and that effectiveness correlates with the length of the programme and with the age of participants - higher effect sizes are associated with longer programmes and older participants. Bowen and Neill (2013) found a similar relationship with participants' ages. Research in this field is limited by the wide range of interventions that can be included with the label 'outdoor adventure learning'. Programmes vary in length, in kinds of activity, in level of challenge, in the ages of participants, and in the inclusion of a residential component, among other factors. This evaluation will test a 5-day intervention, commonly used by schools in the UK.

Non-cognitive outcomes

Many studies have provided evidence for the efficacy of outdoor education in the development of responsibility, leadership development, self-reliance and self-awareness (Bobilya, Klalisch, & Daniel, 2011). Other studies have shown outdoor education as being effective in developing a sense of resilience, a concept that includes perseverance, self-awareness, social support, confidence, and responsibility to others. The working assumption is that increased levels of resilience represent a protective factor, supporting learners in their educational journeys (Ewert & Yoshino, 2011). Further, many meta analyses of outdoor education have pointed to the largely positive impact on young people's attitudes, beliefs and self-perceptions (including self-concept, confidence, self-esteem, locus of control and coping strategies) and interpersonal skills (including communication skills and teamwork). However, reasons why outdoor education works in improving such non-cognitive skills are not fully clear (Hattie, Marsh, Neill, & Richards, 1997; Martin & Leberman, 2005).

Programmes that provide longer, more sustained experiences, appropriate scaffolding and reviewing that facilitates learning are deemed to be more effective (Rickinson, et al., 2004). Meta analyses confirm the notion that outdoor education has positive benefits on children and young people's fitness, motor skills, self-confidence, self-esteem, and relationship with adults, and this finds widespread accord in the literature (Fiennes, et al., 2015). A particular type of outdoor learning - adventure or wilderness therapy, found predominantly outside the UK - claims to offer successful clinical interventions with older young people, families and adults, and to have positive outcomes in terms of self-concept (Bowen & Neill, 2013). There is some recognition in the literature that many of the concepts outlined above are imprecise and definitions vary from study to study, making the study of their development and that of any non-cognitive outcomes a difficult process (Leather, 2013). This has also made analysis of outcomes across studies challenging, as different measures have been used.

Rationale for this study

This evaluation has been designed to address several limitations in the literature. It compares two adventure learning interventions, each taking place over a continuous 5-day period. The study includes a relatively large number of participants in order to ensure sufficient statistical power to test hypotheses. A key feature of this evaluation study is that it will test the effects of the interventions on multiple outcomes, and will investigate relationships among these outcomes over time following the intervention. This will allow us to test some of the findings generated from previous studies and meta-analyses, claiming that adventure learning programmes can have long-lasting effects post-intervention. The design of the evaluation will also provide evidence regarding relationships between non-cognitive measures (self-regulation and engagement) and academic outcomes over time, which is likely to have implications for other areas of educational research and intervention development.

Approach to evaluation

Logic models for both programmes are included in Appendix 1 (OB) and Appendix 2 (CJ's). The logic models inform both the impact evaluation and the implementation and process evaluation. From the review of literature, and discussion with programme leads, key components of the programmes have been identified:

- Group leader as role-model
- Physical challenge
- Experience of agency/autonomy
- Teamwork and building relationships
- Reflection on learning
- Being in nature (OB only)
- Residential; being away from home (OB only)

These components are predicted to give rise to improvements in self-regulation. Combined with follow-up activity in schools, they are predicted to give rise to improvements in student engagement and in turn, attainment. A recent review of evidence showed consistent evidence for associations between non-cognitive skills (including self-regulation and school engagement) and academic outcomes, although this review did not find substantial evidence for causal relationships (Gutman & Schoon, 2013). Gutman and Schoon comment that evidence in this area is limited by the fact that many studies have only included measurement of a single non-cognitive skill, whereas effects are likely to be due to these skills functioning in combination.

The impact evaluation will test for changes in self-regulation, engagement, and attainment. The implementation and process evaluation will explore the relative contributions of different components of each programme, the contribution of school activity post-intervention, and relationships among primary and secondary outcome variables over time.

Impact evaluation

Research questions

1. a. Does an adventure learning intervention[3] lead to changes in non-cognitive skills?

b. Does participation in the Outward Bound intervention lead to different changes in self-regulation of learning[4] compared to a passive control?

c. Does participation in the Commando Joe's intervention lead to different changes in self-regulation of learning[5] compared to a passive control?

d. Do either of these changes persist after approximately 1 year[6]?

2. Does an adventure learning intervention lead to changes in pupil behaviours in schools after approximately 1 year?

3. a. Does an adventure learning intervention lead to changes in mathematics attainment after approximately 1 year?

[3] We use 'adventure learning intervention' here to refer to both the Outward Bound and Commando Joe's programmes. Analyses will compare each of these interventions with the control group, and the two interventions with one another.
[4] dual primary outcome (OB)
[5] dual primary outcome (CJ)
[6] Interventions will take place between September 2019 and January 2020. Immediate post-tests will take place within 2 weeks of the end of the intervention, and follow-up post-tests will take place in October 2020.

b. Does an adventure learning intervention lead to changes after 2 years in general attainment at GCSE (Attainment 8)?

Design

Trial type and number of arms		**Three-arm clustered randomised control trial**
Unit of randomisation		School level
Primary outcome	variable	Pupil self-regulation
	measure (instrument, scale)	Self-Regulation of Learning Self-Report Scale
Secondary outcome(s)	variable(s)	Mathematics attainment after 1 year, General attainment after 2 years; Student engagement, Pupil behaviour in school
	measure(s) (instrument, scale)	GL Progress Test in Maths, GCSE attainment 8 score, The Student Engagement Instrument (SEI), Strength and Difficulties Questionnaire (SDQ; completed by pupil's form tutor)

Randomisation

- Simple randomisation will be employed, without any form of stratification.
- There will be one randomisation block (March 2019).
- This randomisation block will divide schools into OB, CJ and Control group, with equal numbers in each.
- Randomisation will be carried out by a statistician at Sheffield Institute of Education (SIoE) (who is not directly involved in this project) using SPSS. A full syntax audit trail will be published in the final report.

Participants

School eligibility

- Participants will be drawn from secondary schools in England that have a minimum of 20% of pupils eligible for pupil premium.
- Schools will identify 24 current Year 8 pupils who are achieving below their expected levels of progress because of a lack of engagement with education and/or lack of character skills to support learning. At least 50% of participants from each school will be eligible for pupil premium, or be identified by schools as having an equivalent need for support.
- The condition for schools to be entered into the trial (pre-randomisation) will be that the head teacher, chair of governors, the nominated school-based lead and an administrator have signed the Memorandum of Understanding (MoU). Signing of the MoU indicates that opt-in consent has been gained from pupils and their guardians. To be included in the trial, school must also have provided:
    - Names, dates of births and Unique Pupil Numbers (UPNs) for the 24 identified pupils
    - EverFSM status for the 24 identified pupils
    - Details of the 24 pupils' form tutors

Pupil eligibility

At least 50% of the pupils selected should be recognised as disadvantaged, either in receipt of pupil premium or through economic, social or environmental issues known to the school. Schools will be asked to give reasons for selecting pupils as part of the recruitment process.

The programme is especially suitable for pupils with the greatest scope for improvement based on their current attainment levels versus their predicted actual attainment ability, regardless of their starting point.

Sample size calculations

		OVERALL (99 schools)	FSM	OVERALL (105 schools)	FSM
MDES		0.21	0.25	0.21	0.26
School ICC		0.05	0.05	0.05	0.05
Pre-test/ post-test correlations	level 1 (pupil)	0.5	0.5	0.5	0.5
	level 2 (school)	0.25	0.25	0.25	0.25
Alpha		0.025	0.025	0.025	0.025
Power		0.8	0.8	0.8	0.8
Number of schools	OB	33	33	35	35
	CJ's/control	33	33	35	35
	Total per comparison	66	66	70	70
Number of pupils	OB	792 (24 per school)	396 (12 per school)	840 (24 per school)	420 (12 per school)
	CJ's	792 (24 per school)	396 (12 per school)	840(24 per school)	420 (12 per school)
	Control	792(24 per school)	396 (12 per school)	840 (24 per school)	420 (12 per school)
	Total	2376	1188	2520	1260

- Please see Appendix 3 for details on how these MDES estimates were calculated.
- The Alpha value is specified as 0.025, as the trial involves multiple comparisons. A Bonferroni correction has been used (see Appendix 3). The estimates in the table correct for two comparisons (OB v control & CJ's v control). Appendix 3 gives estimates for three comparison (with OB v. CJ's as an additional comparison, which will be reported as an exploratory analysis).
- Level 1 pre-/post-test correlations are estimated based on test-retest reliability of the primary outcome measure (between .69 and .84 over 4 to 6 weeks; Toering et al., 2012), We have reduced to a more conservative .5 due to the longer duration between pre- and post-test in this trial. Level 2 correlations are estimated as .25 following recommendations from EEF.
- School-level ICC has been estimated as 0.05. There is little information in the literature regarding likely ICC for non-cognitive measures. This level was selected following recommendation from EEF. ICC for non-cognitive measures is anticipated to be lower than that for measures of attainment (typically .1-.2).
- A power analysis for the 2-level RCT, with 66 & 70 schools per comparison and 24 pupils per school has been carried out. Similarly, a power analysis was undertaken for the FSM

subsample (12 pupils per school). The target number of schools for recruitment is 99 (33 per arm), but an additional 6 schools are being invited to take part in baseline testing to avoid the risk of too few schools having completed this testing before randomisation.

- The results of this analysis estimate that, for the main ITT analyses the design will be able to detect an effect size of 0.21 or higher as statistically significant (α<0.025) with a statistical power of 0.80.
- The analyses also estimate that for follow-on analyses of pupils ever classed as FSM, the design will be able to detect an effect size of between 0.24 (35 schools per arm) and 0.25 (33 schools per arm) sds or higher as statistically significant (α<0.025) with a statistical power of 0.80

Outcome measures

Time	Dates	Year group	Measures
t1 (baseline)	February 2019	Y8	Self-regulation, School engagement, Behaviour (SDQ)
	March 2019	Y8	Randomisation
	Sept 2019 - Jan 2020	Y9	Pupils take part in intervention
t2 (immediate post-test)	Sept 2019 - Jan 2020	Y9	Within 2 weeks of the end of the intervention: Self-regulation, School engagement
t3 (delayed post-test)	October 2020	Y10	**Self-regulation (primary outcome)**, School engagement, Behaviour (SDQ), Mathematics attainment (GL Progress test in Maths)
	June 2022	Y11	GCSE Attainment 8, to be obtained from NPD when available

Primary Outcome Measure

- Self-regulation of learning will be assessed at three time points using the Self-Regulation of Learning Scale (SRL-SRS; Toering et al. 2012). The primary outcome measure will be self-regulation at t3, at the beginning of Y10. Using this scale, six aspects of pupils' self-regulation are measured, including planning, self-monitoring, evaluation, reflection, effort, and self-efficacy. For this project we will be using a composite score (average of all scales) as the outcome measure. This will be reported as a dual primary outcome measure, separately for each intervention arm.

Self-regulation is a concept employed by leaders of both intervention programmes, and is reported as an outcome in several studies of adventure learning. The instrument employed in this trial consists of 50 items. Test-retest reliability over a 4-6 week period is between .69 and .84 for the six subscales (Toering et al. 2012).

Secondary outcome measures

- Student engagement will be assessed at three time points using the Student Engagement Instrument (SEI; Appleton et al. 2006). This measure focuses on pupils' engagement with schooling and learning. Using this scale, six aspects are measured; these are: teacher-student relationships; peer support at school; family support for learning; control and relevance of school work; future aspirations and goals and intrinsic motivation.

Engagement with learning and with school is also emphasised by leaders of both interventions. Evidence suggests that adventure learning can improve relationships and teamwork among participants, as discussed above. The instrument employed in this trial consists of 33 items. Internal reliability of the six subscales is between .72 and .88 (Appleton et al., 2006).

Self-regulation and School engagement will be measured via the Qualtrics online survey platform. No changes will be made to the original versions of these instruments. School-based lead teachers will coordinate the 24 pupils in their school to complete these two instruments within the window for data collection. The two instruments together are predicted to take up to twenty minutes to complete.

- Student behaviour will be assessed at two time points using the Strengths and Difficulties Questionnaire (SDQ[7]) teacher version and will be completed by each participant's form tutor or equivalent. Using this scale five aspects are measured: emotional symptoms, conduct problems, hyperactivity/inattention, peer relationship problems and prosocial behaviour.
- Maths attainment in Year 10. A mathematics assessment will be administered at the start of year 10 to assess early impact on attainment. This will be the digital Progress Test in Maths, from GL Assessment. The raw score will be used as the outcome measure, as this removes risks associated with incomplete data for participants' dates of birth. Sample size is large enough that distribution of age across groups will be close to equal, and so using standardised score would not confer additional validity This data will be collected using an online format facilitated by GL progress test in maths.
- GCSE Attainment 8. Finally, attainment will be measured at the end of Year 11 as an Attainment 8 score (DfE, 2016). This is calculated by adding up the points for their eight subject and dividing by 10. This will be obtained from NPD data and matched to the other data collected prior to randomisation using pupil UPN.
- We will include KS2 maths score (KS2_MATSCORE) as a covariate in analyses of mathematics attainment at t3. We will include KS2_MATSCORE, KS2_READSCORE and KS2_GPSSCORE as covariates in analyses of attainment at GCSE (Attainment 8).

Analysis Plan

Primary outcome analysis

The primary outcome measure for this trial is the Self-regulation of Learning self-report scale (SRL-SRS) taken at t3, at the beginning of Y10. An intention to treat (ITT) approach will be taken.

To answer research question 1 '***Does an adventure learning intervention lead to changes in non-cognitive skills?'*** a 2-level linear regression model will be undertaken with pupils clustered in schools. The t3 SRL score will be the outcome variable with the trial arm (OB, CJ's or Control) as the independent variable and baseline SRL as a covariate. Comparisons will take place between each arm of the trial and the control condition, OB x Control, CJ x Control using Hedges' g effect size.

To answer research question ***1. a) 'Do these changes persist after approximately 1 year?'*** The same 2-level model will be run with SRL-SRS (t3, approximately 1 year post-intervention) as the outcome variable, t2 (immediate post-test) SRL-SRS as a covariate and with the trial arm (OB, CJ's or Control) as the independent variable.

Statistical uncertainty will be expressed as standard errors of multilevel model coefficients and use of 95% confidence intervals.

Secondary outcome analysis

The secondary analysis to answer research question **2 *'Does an adventure learning intervention lead to changes in pupil behaviours in schools?'*** will use the end point SDQ scores. A 2-level linear regression model will be undertaken with pupils clustered in schools, with the trial arm (OB, CJ's or Control) as the independent variable and SDQ at baseline as the covariate.

[7] http://www.sdqinfo.com/

Comparisons will take place between each arm of the trial and controls; OB x Control, CJ x Control using Hedges' g effect size.

To answer research question three ***'Does an adventure learning intervention lead to changes in attainment?'*** we will use the raw scores on the GL Progress Test in Maths at t3, approximately 1 year after intervention as the outcome variable, with KS2_MATSCORE as a covariate. Again, the trial arm (OB, CJ's or Control) will be included as an independent variable.

To determine whether effect on attainment is present after 2 years GCSE Attainment 8 scores will be used as the outcome variable with KS2 scores as covariates. Again, the trial arm (OB, CJ's or Control) will be included as an independent variable.

Sub Group Analysis

- Sub group analysis of the pupils eligible for FSM will be carried out for both the primary and secondary analyses using the EverFSM variable from NPD. This will be to test whether any improvements seen are the same for FSM as other pupils.
- Pupils identified by schools as disadvantaged, but not eligible for FSM, will not be analysed as a subgroup. There is unlikely to be sufficient consistency among schools in terms of definitions and selections to warrant consideration of these pupils as a homogeneous group.

Implementation and process evaluation

Research questions

4. To what extent do different adaptations at school-level moderate primary and secondary outcome measures? How are changes in attainment due to the intervention mediated and/or moderated by pupils' non-cognitive skills and pupil behaviours in school?

5. How is the intervention delivered to the two 'intention to treat' (ITT) groups, Outward Bound Trust and Commando Joe's? What are the responses from pupils and staff to their experience of the intervention? Do pupils believe that the intervention has improved their non-cognitive skills?

6. What approaches have schools from the ITT groups implemented throughout Y9 and KS4 to build upon the initial intervention? How and why have these approaches been taken? What are the experiences and responses from pupils and staff?

7. How have schools in the control group used funding from the trial to support pupil learning?

The IPE has been designed with reference to the logic models in Appendix 1 and 2. The research questions for the IPE divide into two main areas. The first sets out to examine the relationship between the primary and secondary outcome variables over time. The logic models give rise to the prediction that both interventions should lead to improvement in pupil self-regulation in the short-term. In the medium-term both interventions, combined with school actions post-intervention, should give rise to improved student engagement. Finally, improved self-regulation and student engagement are predicted to give rise to improved behaviour in school and increased attainment. The qualitative data collection and analysis will provide important insight regarding school actions pre- and post-intervention, and the longitudinal SEM (described below) will help to confirm causal relationships between outcome variables over time.

The second main area of investigation for the IPE is to understand any differences in the effectiveness of the interventions associated with school context, or differences in school actions pre- and post-intervention. With respect to this aim, the IPE will provide data on school actions pre- and post-intervention and will explore teacher and pupil perceptions of the interventions and their effects.

Main data collection and analysis activities of the IPE are described below:

Telephone interviews with strategic leads from the OB and CJ's: Building on the work undertaken between intervention leads and SHU during the inception phase, that included a half day IDEA workshop, ongoing dialogue and co-creation of the MOU and this document; additional telephone interviews will be undertaken with strategic leads at the end of the intervention delivery period to ensure deeper understanding of the interventions, how they were delivered and to permit us to update the respective theory of change diagrams detailed in Appendix A and B if required.

Observation visits of OB residentials and CJ's school delivery: During the inception phase, members of the SHU team observed an OB residential involving a secondary school and a CJ's session being delivered at a primary school as part of the year long programme - to inform understanding and data collection tools. Subsequently three visits each will be made during delivery of OB residentials and school-site delivery by CJ's. The aim of these observations is to collect data regarding how consistently OB and CJ's deliver their programme to different schools and in different settings and to provide informal opportunities for discussion with pupils, teachers and the delivery team. This will help the evaluation team to understand what are core components of the respective interventions, and what components vary between schools and settings.

Initial surveys with all school-based leads (SBL): In addition to including the SDQ secondary outcome measure - the baseline survey undertaken prior to randomisation, will seek to understand details about how pupils were identified and contextual details about the school e.g. access to school grounds, provision of extra-curricular clubs and other residential opportunities etc. The post-test survey will repeat the SDQ measure but will differentiate additional questions depending on which arm of the trial schools have been allocated to. For example, OB/CJ'S surveys will seek to understand school level changes, learning among staff, fidelity and perceptions of effectiveness; whereas surveys for schools in the control group will seek to understand what alternative approaches have been taken to improve outcomes for target pupils.

Follow-up surveys with all school-based leads (SBL): To be undertaken in January/February 2020; we would seek to explore areas such as:

- Pupil/school response to the intervention
- Response to the subsequent strategy for school actions to build on the intervention
- Involvement in school clubs (or outside)
- Involvement in other outdoor adventure learning related activity

30 Telephone interviews with SBLs from OB, CJ's and PSG (10 interviews per arm): Sampling for these interviews will be informed by the previous survey of SBLs, and will aim to explore in more depth views expressed in that survey. For OB/CJ's these would take place by the end of the first year of the schools' involvement (Jan-June 2020). Interviews would seek to understand their views of intervention delivery and their perceptions of how pupils experience it; and to explore how as a school they have sought to build upon the initial external input. Following this, based on the surveys and interviews, we will aim to create typologies of approach which could be incorporated into subsequent data collection tools and which could be incorporated into the SEM analysis. We anticipate that schools will vary in the degree to which they follow-up and build on the intervention on return to school, and in the methods that they employ to do this. We will employ a grounded theory approach to our categorisation of follow-up approaches.

The interviews with control group SBLs would be undertaken later (Late Summer or early Autumn 2020) and would be shorter and focus on understanding how schools have/are planning to spend the financial incentive and what approaches have been undertaken to improve Y9 pupil outcomes.

School case study visits: (10 OB, 10 CJ): These would be purposively sampled following initial analysis of the SBL surveys and pupil/form tutor surveys/telephone interviews with SBLs to ensure broadly representative coverage in terms of size, school type, attendance at different OB centres and typologies of approach.

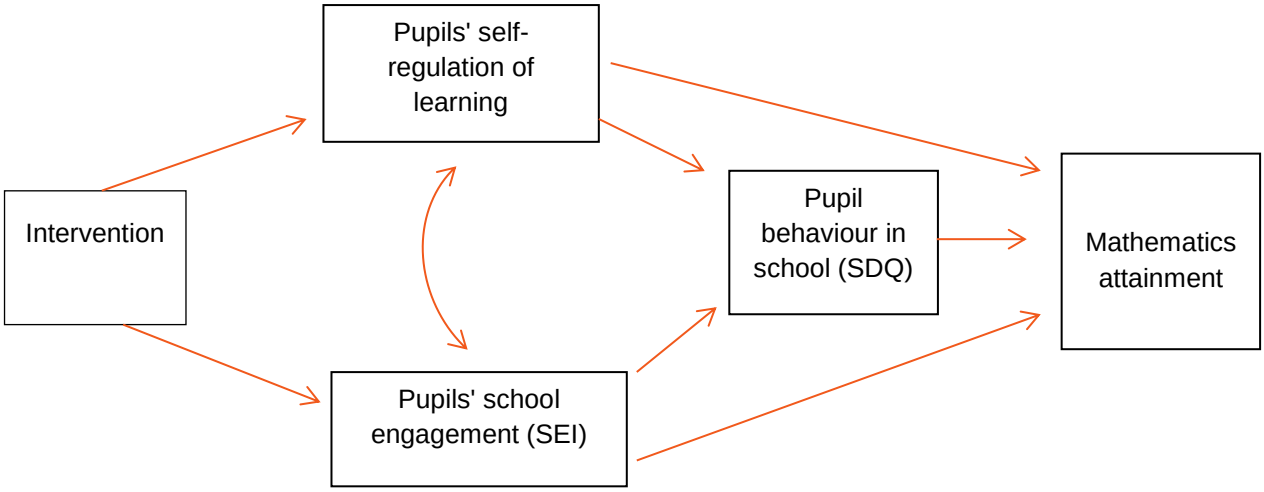
In advance of case study visits, a preliminary telephone interview would be conducted with the SBL to confirm understanding of approaches to implementation being undertaken and to identify the most appropriate individuals to speak with. Although each school is likely to vary, we envisage the following:

- Interview with school-based leads
- Focus group(s) with the Y9 pupils involved
- Focus groups(s) or interviews with relevant staff e.g. pastoral leads, heads of year, SENCO, class teachers and form tutors.

Structural Equation Modelling (SEM)

To complement the qualitative analysis in answering the IPE research questions structural equation modelling will be employed:

- Confirmatory factor analysis (CFA) will be undertaken on each of the non-cognitive measures (SDQ, SRL-SRS, SEI) to check the factor structure.
- Mediation analysis will be conducted to look at whether the effect of adventure learning (X) on attainment (Y) is partially or totally mediated by self-regulation (M1), student engagement (M2) and/or school behaviour (M3). This will be carried for comparison of OB with controls, and of CJs with controls.
  This analysis will be comprised of sets of regression that will test for example:
  Step one: X⇨Y, will test whether adventure learning has an effect on attainment.
  $$Y=b_0+b_1X+e$$
  Step two: X⇨M, this will test whether adventure learning has an effect on self-regulation
  $$M= b_0+b_2X+e$$
  Step three: X+Y=M, will test whether the relationship between adventure learning and attainment is weaker than before or non-significant. If the effect of adventure learning on attainment completely disappears it is fully mediated by self-regulation; if it weakens, it is partially mediated by self-regulation.
  $$Y=b_0+b_4X+b_3M+e$$
- Step four: Will test whether mediation effects are statistically significant using bootstrapping. This will incorporate the longitudinal nature of the SRL and the SEI test measures (not included in the below diagram to prevent complexity)
- Below is a simplified example path diagram that is intended to give an illustration of the intended analysis to be undertaken.
- The straight-headed one-directional arrow represents the 'path' from cause to effect. The double headed curved arrows represent correlations between variables (without causal assumptions).



Cost evaluation

Direct costs will comprise the amount charged by the developers for the 5-day programmes, plus (for groups assigned to the OB group) travel costs to the site.

Teacher time required during the 5-day intervention period will be determined. For OB this will be two teachers for five days. For CJ's, the amount of teacher time may be more variable across schools, and will be determined via the survey of SBLs to be carried out as part of the IPE. There is an initial expectation that two teachers will be involved in the CJ's intervention for the 5 days, but as this intervention takes place on the school site there may be some flexibility in how this is interpreted.

Ethics and registration

The evaluation of the Adventure Learning programme has been given independent ethical approval by the Sheffield Hallam University ethics committee.

Opt-in informed consent will be gained from participating pupils and their guardians. (NB. The legal basis for data processing used by SHU is public task, as specified below.)

The trial will be registered at www.controlled-trials.com once the number of participating schools can be confirmed, and before September 2019 when the intervention begins. The ISRCTN (International Standard Randomised Controlled Trial Number) will be updated within this document when available. The trial registry at www.controlled-trials.com will be updated with the findings of the trial once the project is complete.

Data protection

In accordance with the General Data Protection Regulation, Chapter 2, Article 6 the legal basis for the processing of personal data for this project is 'public task'.

At all points SHU are responsible for retrieving and processing data as part of this trial using password protection and secure transfer methods such as SHU ZendTo.

For the purpose of research, the responses will be linked with information about the pupils from the National Pupil Database and shared with the EEF, EEF's data processor for the archive (currently FFT Education), Department for Education and, in an anonymised form, with the Office for National Statistics[8] and potentially other research teams.

[8] Pending new procedures being finalised by the DfE and ONS.

We will not use pupil/staff names or school names in any report arising from the research.

Further matching to NPD data may take place during subsequent research. For transparency, the precise terms of this data sharing will be stated in a fair processing notice, specifying the personal data (pupil names, pupil ID numbers, FSM status, KS2 attainment, KS4) to be processed, in line with GDPR.

A data sharing agreement will detail the personal data to be shared, and a fair processing notice will be sent to all participating schools as per GDPR requirements.

Personnel

Outward Bound Trust

- Daniel Cibich: Head of education partnerships: To oversee the recruitment of schools for the trial (pre-randomisation stage). To oversee the booking of schools allocated to Outward Bound group onto Outward Bound courses in line with programme timetable. To oversee Outward Bound delivery aspects of the trial in conjunction with relevant operational staff at Outward Bound's residential centres. To liaise with CJs and SHU on all delivery related aspects as relevant.
- Isabel Berry: Head of trusts and foundations: To oversee Outward Bound's project management and liaison through the trial with EEF and, as relevant with CJs and SHU.
- Emma Ferris: Head of evaluation: To advise Outward Bound, as requested, on aspects relating to the delivery and evaluation of the trial.

Commando Joe's

- Michael Hamilton: Founder/director
- Daniel Kelly, : National Manger

Sheffield Hallam University

- Tim Jay: Principal investigator
- Sarah Reaney-Wood: Co- project manager and impact evaluation lead
- Ben Willis: Co- project manager and IPE lead
- Jean Harris-Evans: Researcher
- Bronwen Maxwell: IPE advisor
- Sean Demack: Impact evaluation advisor

Risk analysis

Risk	*Who could be harmed?*	*Mitigation*	*Risk level*
Potentially high levels of attrition due to pupils leaving school	Evaluation	Sample calculations include contingency for school population turnover. Pupils will only be included in analyses if they continue in the same school for the three years of the project.	Low
Pupils will be spread across different classes within each school and so may be difficult to	Evaluation	As part of the recruitment process, schools will be required to nominate a lead teacher (likely to be a teacher who will accompany the group during the intervention). The lead teacher will be responsible for arranging data collection (primarily survey	Low

track or contact for data collection		completion) over the two years following the intervention.	
Not possible to control pupils' involvement in other extra-curricular activity	Evaluation	It would be unethical to stipulate that participating pupils should not be involved in any residential or other enrichment activity outside of the project. However, we will request that schools in control group do not offer residential outdoor adventure learning to target pupils. We will also ensure that all enrichment activity that participating pupils engage with is recorded by the lead teacher in each school so that this data can be included in analyses. Schools that have previously offered Outward Bound courses to the target cohort are excluded from the trial.	Medium
Data management and GDPR	Pupils/School staff	A joint Outward Bound /CJ / SHU opt-in consent process will be designed for schools, parents and pupils for both intervention and data collection. The team has substantial experience in negotiating ethics processes and will be supported by the Faculty ethics committee in this. Informed opt-in consent will satisfy requirements of GDPR legislation for data storage and use for the project.	Low
Schools may have different criteria for recording behavioural transgressions	Evaluation	We will request numbers of detentions and exclusions as a measure of pupil behaviour in Year 10 and 11. Schools may have different criteria for these, and may have different systems of recording. We will assess validity of these measures as part of our process evaluation and will consider removing these from analyses if validity is below expectations.	Low
Recruitment may be challenging if we include a passive control	Evaluation	Schools in the control group will be offered funding equivalent to the cost of the active control conditions to use to enhance the learning of their target group. Part of the process evaluation for the project will involve the monitoring of the use of these funds, plus all curriculum enrichment activity across the three conditions.	Low
Longitudinal data collection will be challenging, with 99 schools involved	Evaluation	Ensure sufficient admin support within project team to allow time to chase up lead teachers to respond. Stipulate commitment to complete surveys in memorandum of understanding with schools.	Low
Misalignment of CJ and OB intervention means that comparison could be difficult	Evaluation	CJ & OB should be as closely aligned as possible. Both interventions will take place over a continuous 5-day period.	Low
Observations taking place in remote areas	Evaluation team	All of the evaluation team will follow health and safety procedures during all observation and will take measures (e.g. sensible clothing and footwear) to minimise risk	Low

Timeline

December 2018	Schools to return 24 signed pupil/guardian consent forms to Outward Bound. Memorandum of Understanding to be signed by 1) Headteacher, 2) School based lead, 3) School administrator and 4) Chair of governors and returned to OB.
February 2019	Baseline data collection for OB, CJ and Control groups: pupils: Self-regulation, School engagement SBL: 1st SBL survey pupils' form tutors: SDQ
March 2019	Schools randomised to one of the three groups
September 2019 - January 2020	CJ/OB delivery *(3 CJ and 3 OB schools will be selected by the evaluation team, in consultations with intervention leads for observations during delivery)*
September 2019 - January 2020	T2 data collection for OB/CJ groups (*to take place within two weeks of receiving the intervention*) pupils: Self-regulation, School engagement
January-February 2020	T2 questionnaire for control schools to take place (as above)
January 2020	2nd survey of SBLs to take place across ALL schools
March 2020-July 2020	SBL telephone interviews for CJ and OB schools, sub-sample of 10 each.
May 2020-December 2020	Case study visits will take place in a sub-sample of CJ/OB schools (10 in each)
June - October 2020	SBL telephone interviews for control schools in a sub-sample of 10
October 2020	T3 data collection in ALL schools pupils: Self-regulation, School engagement, GL Progress Test in Maths pupils' form tutors: SDQ
March 2021	Report first draft submitted to the EEF
Autumn 2022	GCSE results available (NPD)
March 2023	Report addendum submitted (long-term follow-up: GCSE Attainment 8)

References

Appleton, J. J., Christenson, S. L., Kim, D., & Reschly, A. L. (2006). Measuring cognitive and psychological engagement: Validation of the Student Engagement Instrument. *Journal of School Psychology*, *44*(5), 427-445.

Betts, J. E., Appleton, J. J., Reschly, A. L., Christenson, S. L., & Huebner, E. S. (2010). A study of the factorial invariance of the Student Engagement Instrument (SEI): Results from middle and high school students. *School Psychology Quarterly, 25.*

Bloom, H.S. (1995). Minimum Detectable Effects: a simple way to report the statistical power of experimental designs. *Evaluation Review 19*(5), 547-556.

Bloom, H.S. (2006). *The Core Analytics of Randomized Experiments of Social Research*. MDRC Working Papers on Research Methodology, August 2006. Available at https://www.mdrc.org/publication/core-analytics-randomized-experiments-social-research

Bloom, H, S., Richburg-Hayes, L. & Black, A. R. (2007). Using Covariates to Improve Precision for Studies That Randomize Schools to Evaluate Educational Interventions. *Educational Evaluation and Policy Analysis 29*(1), 30–59

Bobilya, A., Klalisch, K., & Daniel, B. (2011). *Journal of Experiential Education, 33*(4), 356–359.

Bowen, D., & Neill, J. (2013). A Meta-Analysis of Adventure Therapy Outcomes and Moderators. *The Open Psychology Journal, 6*, 28-53.

Cason, D., & Gillis, H. L. (1994). A meta-analysis of outdoor adventure programming with adolescents. *Journal of Experiential Education, 17*(1), 40-47.

Department of Education. (2016). *Progress 8: How Progress 8 and Attainment 8 measures are calculated*. London: Department of Education.

Ewert, A., & Yoshino, A. (2011). The influence of short-term adventure-based experiences on levels of resilience. *Journal of Education and Outdoor Learning, 11*(1), 35-50.

Fiennes, C., Oliver, E., Dickson, K., Escobar, D., Romans, A., & Oliver, S. (2015). *The Existing Evidence-Base about the Effectiveness of Outdoor Learning.* London: Giving Evidence (in partnership with UCL, Institute for Outdoor Learning, The Blagrave Trust).

Gutman, L., & Schoon, I. (2013). *The impact of non-cognitive skills on outcomes for young people*. London: Education Endowment Foundation.

Hattie, J., Marsh, H., Neill, J., & Richards, G. (1997). Adventure Education and Outward Bound: Out-of-Class Experiences That Make a Lasting Difference. *Review of Educational Research, 67*(1), 43-87.

Hedges, L.V. & Rhoads, C. (2010). *Statistical Power Analysis in Education Research*. NCSER 2010-3006. Available at https://ies.ed.gov/ncser/pubs/20103006/pdf/20103006.pdf

Leather, M. (2013). 'It's good for their self-esteem': the substance beneath the label. *Journal of Adventure Education & Outdoor Learning, 13*(2), 158–179.

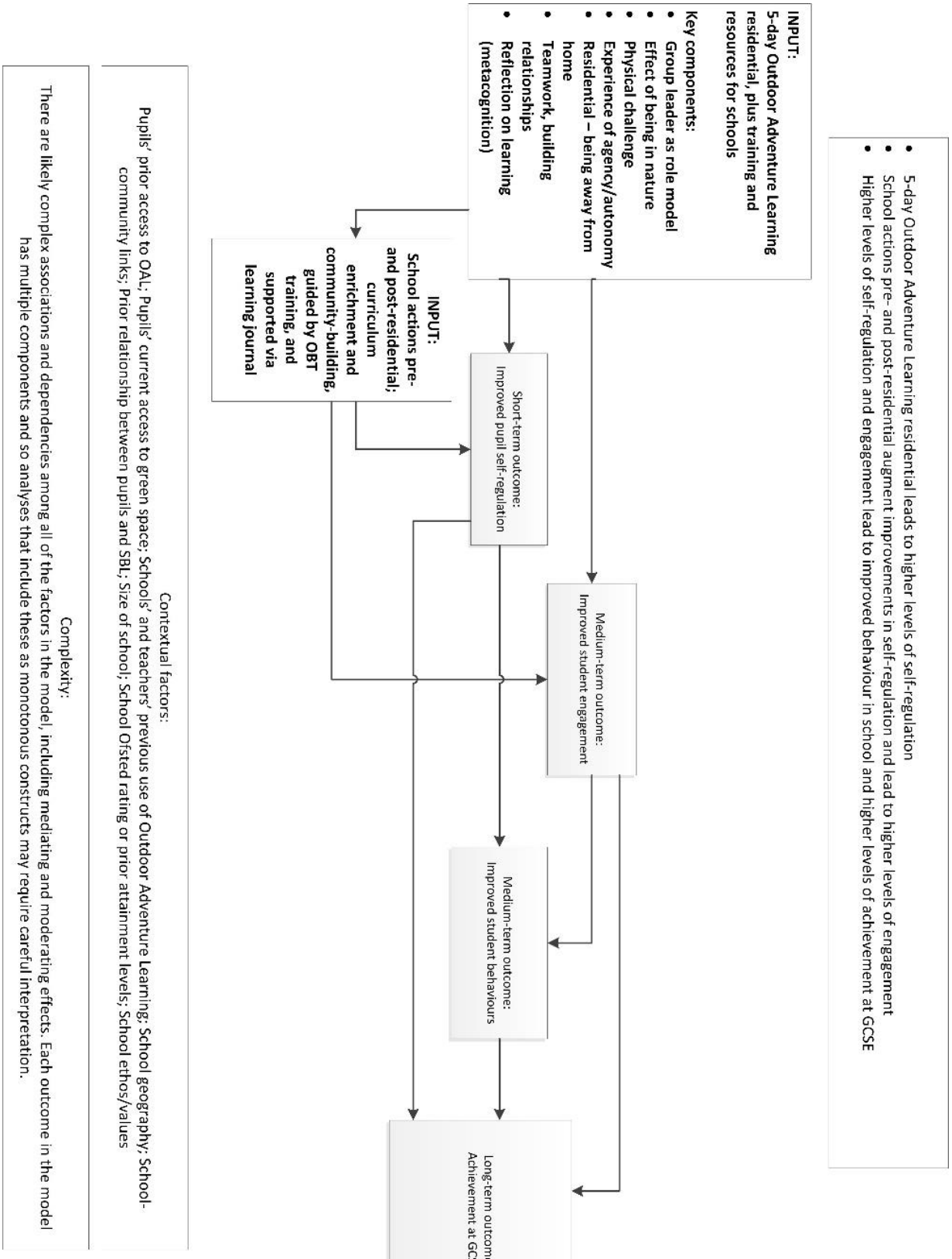
Martin, A., & Leberman, S. (2005). Personal Learning or Prescribed Educational Outcomes: A Case Study of the Outward Bound Experience. *Joumal of Experiential Education , 28*(1), 44-59.

Rickinson, M., Dillon, J., Teamey, K., Morris, M., Young Choi, M., Sanders, D., & Benefield, P. (2004). *A Review of Research on Outdoor Learning*. London: Field Studies Council.

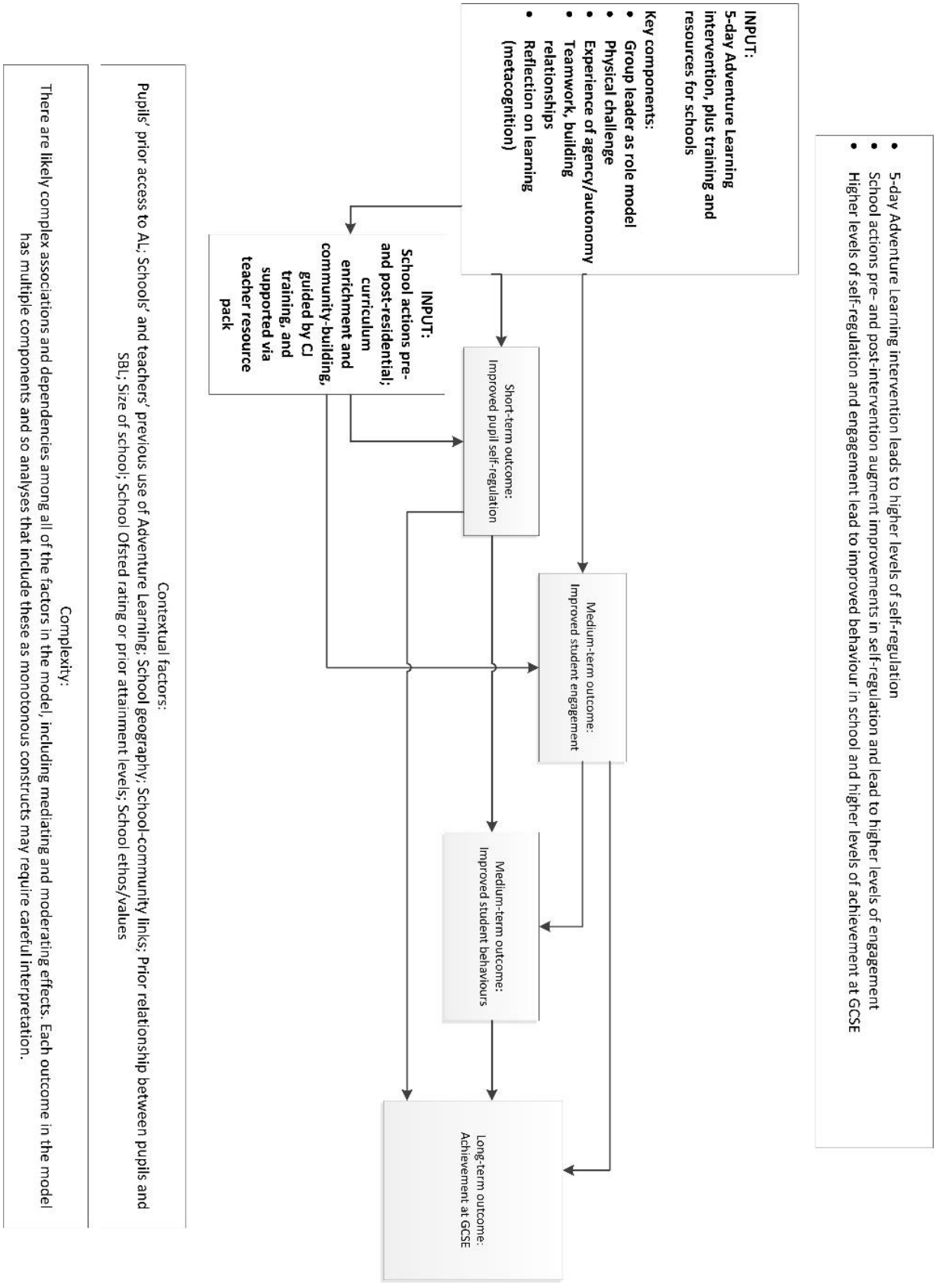
Toering, T., Elferink-Gemser, M. T., Jonker, L., van Heuvelen, M. J., & Visscher, C. (2012). Measuring self-regulation in a learning context: Reliability and validity of the Self-Regulation of Learning Self-Report Scale (SRL-SRS). *International Journal of Sport and Exercise Psychology, 10(*1), 24-38.

Wason, J. M., Stecher, L., & Mander, A. P. (2014). Correcting for multiple-testing in multi-arm trials: is it necessary and is it done? *Trials*, *15*(1), 364.

Appendix 1: Logic Model, OB



Appendix 2: Logic model, CJ



Appendix 3: Power Analyses

From Bloom et al. (2007), the Minimum Detectable Effect Size (MDES) for a 2-level CRT is:

$$MDES_{2LCRT} = M_{K-L-2}\sqrt{\left(\frac{ICC_{Sch}(1-R^2_{Sch})}{\mathrm{P(1-P)K}}\right)+\left(\frac{(1-ICC_{Sch})(1-R^2_{pup})}{\mathrm{P(1-P)Km}}\right)}$$

Where:

P is the proportion of schools who receive the intervention (=0.50)
R^2_{pup} is the pupil-level covariate explanatory power (=0.50x0.50=0.25)
R^2_{Sch} is the school-level covariate explanatory power (=0.25x0.25=0.0625)
ICC_{Sch} is the cluster (school) level Intra Cluster Correlation coefficient (=0.05)
K is the number of clusters (schools =66 or 70)
Km is the total number of individual (pupils=792x2=1,584; 840x2=1,680)
m is the number of pupils per school (=24)
L is the number of cluster level covariates (=2)
M_{K-L-2} is the t-distribution multiplier with K-L-2 degrees of freedom - which equates to between 62 (with 33 schools per arm) and 66 (35 schools per arm) degrees of freedom.

1:1 or multiple comparisons.
A 1:1 comparison involves (for example) comparing the OB group with the control group. This ignores the increased risk of type I errors associated with multiple comparisons (Wason et al., 2014). In this trial there are 3 intervention groups; the intervention (OB); an active control (CJ's) and a business-as-usual control. Therefore, there will be up to 3 comparisons:

- OB v Control
- CJ's v Control
- OB v CJ's

The first two of these relate to testing whether the two interventions have an impact relative to the business as usual control group. The third might be seen as a follow-on analysis that tests whether a significant difference exists between the two interventions. To control for the inflation of type I errors associated with multiple comparisons, a Bonferroni correction is applied by dividing the probability of making a type I error (α) by the number of proposed comparisons.

The power analyses now proceeds in three stages, first assuming a 1:1 comparison, second assuming two comparisons (OB v control & CJ's v control) and finally assuming three comparisons (OB v CJ's; OB v control; CJ's v control).

NOTE - for the headline ITT analyses, two comparisons are assumed - represented a comparison of each intervention with the control group..

Assuming 1:1 comparison and a 2-tailed test; α=0.05, α/2=0.025; statistical power of (1-β=0.80). M_{62} = 2.8464. ; M_{66} = 2.8437.

Therefore, assuming 33 schools per arm (66 in each comparison);

$$MDES_{2LCRT} = 2.8464\sqrt{0.00464} = 0.1939 \sim 0.19$$

Assuming 35 schools per arm (70 in each comparison);

$$MDES_{2LCRT} = 2.8437\sqrt{0.00438} = 0.1881 \sim 0.19$$

So, if multiple comparisons are ignored, the MDES estimate is 0.19 standard deviations.

Assuming two comparisons (OB v control; CJ's v control) and a 2-tailed test; Bonferroni correction used for α, α per comparison = α/2 = 0.025, two-tailed = 0.0125; Statistical power of (1-β=0.80).
M_{62} = 3.1446. ; M_{66} = 3.1408.

Therefore, assuming 33 schools per arm (66 in each comparison);

$$MDES_{2LCRT} = 3.1446\sqrt{0.00464} = 0.214 \sim 0.21$$

Assuming 35 schools per arm (70 in each comparison);

$$MDES_{2LCRT} = 3.1408\sqrt{0.00438} = 0.208 \sim 0.21$$

Correcting for 2 comparisons increases the MDES estimate to 0.21 sds.

Assuming three comparisons (OB v CJ's; OB v control; CJ's v control) and a 2-tailed test; Bonferroni correction used for α, α per comparison = α/3 = 0.0167, two-tailed = 0.0083'; Statistical power of (1-β=0.80). M_{62} = 3.1446.
M_{62} = 3.3081. ; M_{66} = 3.3036.

Therefore, assuming 33 schools per arm (66 in each comparison);

$$MDES_{2LCRT} = 3.3081\sqrt{0.00464} = 0.225 \sim 0.23$$

Assuming 35 schools per arm (70 in each comparison);

$$MDES_{2LCRT} = 3.3036\sqrt{0.00438} = 0.218 \sim 0.22$$

Correcting for 2 comparisons increases the MDES estimate to between 0.22 and 0.23 sds.

For the planned follow-on subgroup analyses involving pupils ever classed as FSM, the same equation is used but with a reduced number of FSM pupils per school:

Km is the total number of individual (FSM pupils=396x2=792; 420x2=840)
m is the number of pupils per school (=12)

- Assuming 1:1 comparison: MDES estimate = 0.22 (70 schools) - 0.23 (66 schools)
- Correcting for 2 comparisons: MDES estimate = 0.24 (70 schools) - 0.25 (66 schools)
- Correcting for 3 comparisons: MDES estimate = 0.26 (70 schools) - 0.27 (66 schools)

Summary

For the headline ITT analyses we estimate that the proposed design will be able to detect an effect size of 0.21 sds for all pupils and an effect size of between 0.24 and 0.25 for FSM pupils.