Adventure Learning Trial

Longitudinal and Addendum Analysis

Date agreed with EEF:
Developer: The Outward Bound Trust and Commando Joe's
Evaluator: Sheffield Hallam University
Principal Investigators: Dr Sarah Reaney-Wood & Benjamin Willis
Statistician: Dr Sarah Reaney-Wood & Sean Demack

Introduction

Adventure learning (the programme) is evaluating whether a 1-week adventure learning experience leads to increases in pupil self-regulation, student engagement, reducing behavioural difficulties and improving prosocial behaviour in school, and general attainment, either directly or indirectly. In an efficacy trial, the team investigated whether adventure learning-type activities lead to an improvement in self-regulation, student engagement, and behaviour among hard-to-reach students in Year 9 (age 13 to 14). Specifically, the trial tested two approaches to adventure learning, comparing a residential programme provided by Outward Bound Trust (OBT) to a control group and then comparing a programme conducted on school grounds provided by Commando Joe's (CJs) to a control group to examine whether a residential element provides the catalyst for this change over a school-based programme.

As a follow up to the 2023 published efficacy trial, the team aims to conduct longitudinal analysis to examine the potential impact on GCSE scores (using Attainment 8 data) for the students that took part in the trial, when they reach Year 11. In the first set of analysis presented in the 2023 report the effect of Adventure Learning on self-regulation, student engagement and behaviour was explored, this longitudinal analysis follows directly from this, now that their GCSE outcomes are available. The aim of the longitudinal analysis is two-fold. Firstly, to look at the impact of both the Outward Bound Trust (OBT) and Commando Joe's (CJs) programmes on academic attainment, in the students that received the interventions. This will require access to attainment 8 data for the students that took part in the trial, using the National Pupil Database (NPD). Attainment 8 data is a measure of general attainment and is calculated by adding up the points for their 8 subjects, with English and Maths counted twice (Department of Education, 2016). Secondly, the longitudinal analysis will aim to look at the relationships between the psychological outcomes of self-regulation and student engagement, behaviour and attainment outcomes, and to explore and test theoretical assumptions that the impact of adventure learning on attainment is mediated by self-regulation, student engagement and improved behaviour.

The purpose of this statistical analysis plan is to provide detail on the longitudinal analysis to be conducted on this sample and it follows the guidance published by EEF for longitudinal analysis on trials. Detailed descriptions of related literature, the main trial design, sample size calculations, etc. are available in the statistical analysis plan (SAP) and the report of the main trial.

Research questions

The addendum analysis aims to answer three research questions that could not be answered during the main trial analysis due to longitudinal nature of the data. RQ1 continues from the main trial impact analysis. RQ2 and RQ3 pertain to mediation analysis.

Impact analysis RQs:

Research Question 1 (RQ1): Does an adventure learning intervention lead to changes in general attainment at GCSE (attainment 8) two years after programme delivery?

a. What is the impact of the OBT programme on attainment compared to students who do not receive any adventure learning programme (OBT vs. control) two years after programme delivery, as measured by Attainment 8 scores?
b. What is the impact of the CJ programme on attainment compared to students who do not receive any adventure learning programme (CJ vs. control) two years after programme delivery, as measured by Attainment 8 scores?
c. What is impact of the OBT programme on attainment compared to the CJ programme (OBT vs CJ) two years after programme delivery, as measured by Attainment 8 scores?

Correlational and mediation analysis RQs:

Research Question 2 (RQ2): What relationships exist between the intervention, three non-cognitive outcomes - self-regulation, student engagement, behaviour in school and attainment?

Research Question 3 (RQ3): How are any changes in general attainment among the students who receive an adventure learning intervention mediated by students' self-regulation, student engagement and behaviour in school?

Table 1: Relevant links between main trial outcomes and outcome measures for longitudinal analysis

Trial design, including number of arms		Three-arm clustered randomised controlled trial
Unit of randomisation		School
Stratification variables (if applicable)		N/A (simple randomisation used)
Primary outcome	variable	Self-regulation after 1 year (this differs between schools due to the impact of COVID 19 to between 12 and 17 months)
	measure (instrument, scale, source)	Self-regulation of Learning Self-Report Scale (SRL-SRS) (Toering, 2012)
Secondary outcome(s)	variable(s)	Self-regulation immediate post-test; Student engagement (immediate post-test and after 1 year); Pupil behaviour in school after 1 year approximately (this was delayed due to COVID 19 school closures)
	measure(s) (instrument, scale, source)	SRL-SRS; Student Engagement Instrument (SEI); Strengths and Difficulties Questionnaire (SDQ)

Longitudinal outcome	variable(s)	GSCE General attainment after 2 years;
	measure(s) (instrument, scale, source)	Attainment 8 (NPD);

Outcome measures

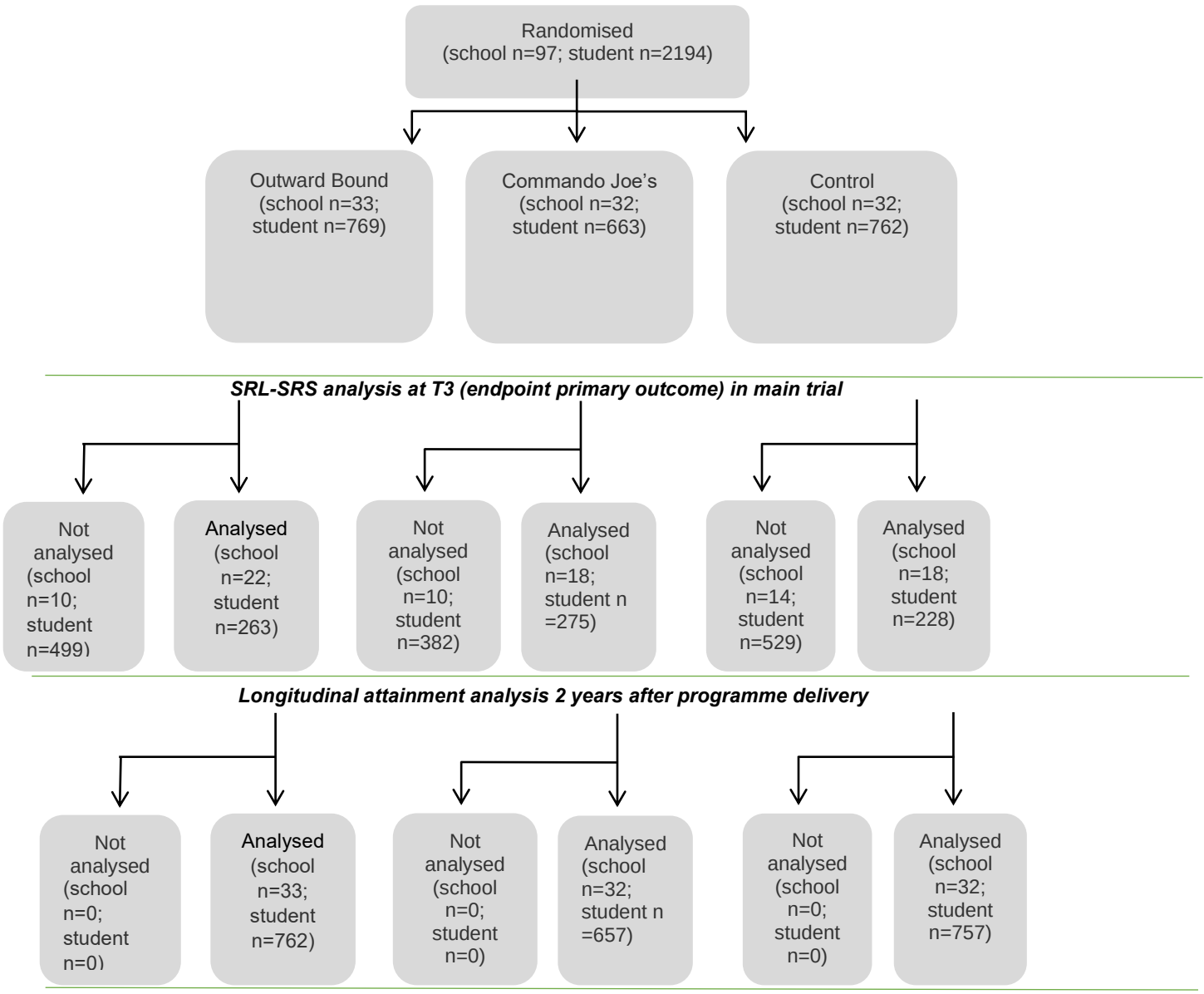
Table 2: Outcome measures for longitudinal analysis

Outcome	Measurement tool	Description
Self-regulation of Learning immediately post intervention and longitudinal follow-up	Self-regulation of learning self-report scale (SRL-SRS) (Toering et al., 2012)	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.
Student Engagement immediately post intervention and longitudinal follow-up	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. As with the main trial analysis, we will be using a composite score (average of all scales) as the outcome measure.
Behaviour Longitudinal follow-up	Strengths and Difficulties Questionnaire (SDQ) (Goodman., 2001)	Using this scale five aspects are measured: emotional symptoms (ES), conduct problems (CP), hyperactivity/inattention (HI), peer relationship problems (PP) and prosocial behaviour. Internalising (sum of ES and PP) and externalising (sum of CP and HI) representing behavioural difficulties and the prosocial score, representing strengths will be used.
Attainment	Attainment 8 score	Ranges from 1 (lowest) to 9 (highest) and is calculated by adding up the points for their 8 subjects, with English and Maths counted twice (DfE) (Average grade) (Department of Education, 2016).

Participants

We requested NPD data for pupils in the main trial sample. A total of 2194 pupils are part of the sample. We requested data on 2194 pupils.

Figure 1: Participant flow diagram (three arms)



Access to NPD

We applied to the NPD for individual level Attainment 8 data. We provided pupil data (Name, UPN, D.O.B, School name and unique ID) for the DfE to match the requested NPD data. The export of NPD data will be transferred to the Office of National Studies (ONS) for the

Evaluation Team to analyse and will not contain personal data. Access to NPD data is via the ONS Secure Research Service. The trial statistician Dr Sarah Reaney-Wood is an ONS accredited researcher who submitted the application and will access and analyse the data when authorised.

Analysis plan

All of the analysis detailed within this SAP should be considered exploratory and it deals with longitudinal analysis of secondary outcomes and the relationships between variables, for which the trial was not powered for. The analysis to answer RQ1 will follow an Intention to Treat (ITT) approach. Outcome data will be summarised descriptively for the three groups - OBT, CJ and control (Example Table 3). The same significance levels used in the main trial will be used. The outcome for RQ1 is general attainment (as measured by attainment 8 scores, with the first model measuring the impact of the OBT programme on attainment compared with the business as usual (BAU) control group (RQ1a: OBT v control) and the second measuring the impact of the Commando Joe programme on general attainment compared with the business-as-usual (BAU) control group (RQ1b: CJs v control). A third model will be used to compare the impact of the OBT and CJ programmes (OBT v CJ) on attainment. A pupil-level key stage 2 covariate taking an average of maths, reading and grammar punctuation and spelling (GPS) will be included in all models.

Effect size calculations based on the difference between the groups two years after programme delivery will follow the same approach as in the main trial analysis and will be presented as Hedges' g with 95% confidence intervals (CI) (Example Table 4). The intra-cluster correlation coefficient (ICC) two years after programme delivery will be reported at the school level.

Table 3: Descriptive statistics of attainment outcome by trial arm

Table 3 descriptive statistics of attainment outcome by trial arm	Longitudinal attainment					
	OB (n)		CJ (n)		Control (n)	
	Mean	Sd	Mean	Sd	Mean	Sd
General attainment						
KS2 average						

Table 4: Example (empty) output data

	Unadjusted means				Effect size		
	Intervention group		Control group				
Outcome	n (missing)	Mean (97.5%/95% CI)	n (missing)	Mean (97.5%/95% CI)	Total n (intervention; control)	Hedges g (97.5%/95% CI) *	p-value
General attainment (attainment 8) OBT vs Control							
General attainment (attainment 8) CJ vs Control							
General attainment (attainment 8) OB vs CJ							

Subgroup Analysis

As with the main trial analyses, subgroup analysis will be conducted for FSM pupils using the same multi-level linear model approach as described above for RQ 1, which is the same as in the main trial. The FSM subgroup will be identified by the EverFSM_6_P indicator from the NPD data set.

Descriptive statistics will be used to describe the portion of the sample that were identified as FSM. As with all the other analysis within the SAP, this analysis is exploratory as the trial was not powered to detect effects on subgroups.

Correlation and Mediation Analysis

Research question 2 'What relationships exist between the intervention, three non-cognitive outcomes - self-regulation, student engagement, behaviour in school and attainment?'

Correlation analysis will be conducted to explore the relationships between all variables; intervention status, self-regulation, student engagement, SDQ internalising score, SDQ externalising score, SDQ prosocial score and attainment. For each pairing we will report the strength and direction of the relationship between the two variables and a p-value. The type of correlational analysis undertaken will depend on the variables in question. For example, Point biserial when including dichotomous variables (Intervention) and Pearson correlation for all other variables.

Exploration of the association between the above outcomes/variables are supported by theory. However, this analysis is being conducted to help confirm the theoretical assumptions of associations between the intervention and the outcomes, drawing on the programmes Theory of Change and to support the inclusion of each variable in the subsequent mediation analysis outlined below. Specifically, where there is no association between the intervention and the mediating variables, these mediators will not be included in the model. This approach is considered appropriate due to the exploratory nature of this analysis, the potential complexity

of the model if all variables are included without prior correlation analysis (which increases the risk of over specifying the model) and the limitation of sample size.

Each of the outcomes will be scored according to their protocol, leaving a total score for self-regulation self-report scale at two timepoints, student engagement instrument at two timepoints, and a internalising score, externalising score and prosocial behaviour for the strength and difficulties questionnaire at longitudinal follow-up only. As in the main analysis conducted in the recent report, Willis et al., (2023) centring will not be used on the variables.

Research question 3 'How are any changes in general attainment among the pupils who receive an adventure learning intervention mediated by pupils' non-cognitive skills and pupil behaviours in school?' focuses on understanding whether the impact of adventure learning on attainment is mediated by self-regulation, student engagement and behaviour. 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 & M2), student engagement (M3 & M4) and/or school behaviour (M5, M6 & M7). This will be carried out for comparison of OB with controls, and of CJs with controls. Analysis from research question two above is driving the inclusion of variables into the mediation analysis. Variables available for inclusion were outcomes in the trial based on underlying theoretical assumptions on the relationships between adventure learning, improvement in psychological outcomes and attainment.

Mediation analysis will include data with multiple time points (SRL-SRS, SEI) and data collected at just one follow-up up point (SDQ), requiring a model that includes both parallel and sequential mediators. Through this, we will explore whether improvements (including longer-term improvements) in self-regulation, SEI and behaviour are needed to see changes in attainment.

Table 5: Variables to be included in analysis

Variable type	Description
Independent variable	Adventure learning involvement
Dependent variable	Attainment measured using attainment 8
Mediators	Self-regulation of Learning total score measured at immediate post-test (T2) (M1) and at longitudinal follow-up (T3) (M2)
	Student engagement total score measured at immediate post-test (T2) (M3) and at longitudinal follow-up (T3) (M4)
	Student externalising behaviour (M5), internalising behaviour (M6) and prosocial behaviour (M7) measured at longitudinal follow-up (T3) only

Rational for the choice of mediators

Previous evidence suggests that outdoor adventure learning can lead to raised attainment (Cason & Gills, 1994; Hattie et al., 1997), but the mechanisms underpinning this are unclear. Similarly, outdoor education has been suggested to develop non-cognitive/psychological skills such as; responsibility, leadership development, self-reliance, and self-awareness (Bobilya et al., 2011), resilience (a concept that includes perseverance, self-awareness, social support, confidence, and responsibility to others) and have a 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).

Self-regulation of learning, Student engagement and behaviour (internalising, externalising and prosocial behaviour) were selected as the outcomes in the trial and as the mediators for consideration in mediation analysis based on our theoretical and previous evidence that suggests that improvements in these non-cognitive outcomes may be the mechanism by which an increase in attainment is seen. The literature provides evidence for improvements in both attainment and non-cognitive factors following participation in outdoor adventure programmes and highlights the need to carry out research studies that explore interactions among measures over time, as well as looking for changes in individual measures.

Model specifications

Using the Lavaan package in R, the direct effects, indirect effects, and total effects can all be explored within one model. Essentially, this approach comprises of sets of regressions that will test for partial or full mediation.

Consideration will be given to the possibility of any measured confounding variables that may explain the relation between any two variables. From the main trial analysis we do not anticipate any measured confounders, however, should any confounder be related to either predictor or mediator, it will be accounted for in the analysis.

As we have the potential (dependent on the outcome of research question two) of both parallel and sequential/serial mediators, the model is slightly more complex. Furthermore, from the literature, the understanding of the longitudinal effects of adventure learning on attainment and psychological characteristics is lacking. As such, priority will be given in this model to understand each of the possible mediating paths, both between immediate post-test improvements in regulation and engagement and longitudinal improvements in regulation, engagement and behaviour on attainment.

Effects of interest

Direct effects of adventure learning on attainment:

Adventure Learning – Attainment:

Indirect effects of adventure learning on attainment:

Following the correlational analysis specified for research question two above, we will explore the indirect effects of adventure learning on attainment through the mediators. Only mediators that have an association with the intervention will be included in the models.

An example of an indirect path we would expect to explore is:

Example path**:** Adventure Learning-self regulation T2- student engagement at T2 -attainment (X-M1-M3 -Y)

This path represents a joint mediator model where self-regulation and student engagement mediate the relationships between the intervention and increased attainment.

Total effect:

We will explore the total effects

Direct + Indirect effects

The mediation effect will be compared to the direct effects to see whether the relationship is fully or partially mediated.

Fit indices will be used to determine how well the specified model fits the data. Some of the commonly used fit indices within mediation analysis using Lavaan are model chi-square (X^2), comparative fit index (CFI) , root mean square error of approximation (RMSEA), Tucker Lewis index (TLI) (Hooper, Coughlan & Mullen., 2008; Ballen & Salehi., 2021). We will use the following as indicators of good fit:

- model chi-square (X^2) insignificant result at a 0.05 threshold

- Comparative fit index (CFI) ≥ 0.95

- Root mean square error of approximation (RMSEA) of 0.08 and below as good fit

In the case of poor model fit, we will closely examine the structure of the defined model, including non-contributing paths, and consider the theoretical basis for the inclusion/exclusion of paths. We will then adjust the model and compare the fit of the new model against the older model. However, only paths with previous theoretical justification will be considered for model improvement (Saris, Satorra, Van der Veld., 2009). Robustness indices will be requested using the 'robust' variants function within the Lavaan package.

Mediation will be supported by the use of bootstrap confidence intervals, as this can be the more powerful and pragmatic approach to use (Hayes & Scharkow., 2013), improving our confidence in the findings.

Given the potential complexity of mediation analysis and especially as there are multiple mediators, as well as reporting the findings using statistical outputs and tables, we will include a graphical visualisation of the model including the estimated coefficient (b), the standard error of the coefficient (SE) and the p value of significant (p) for each path (Ballen & Salehi., 2021).

Limitations

Limitations are related to both the last element of the impact analysis and the mediation analysis.

Impact analysis

Attrition across the course of the trial was high, however, the level of attrition for the attainment outcomes is expected to be less as this only requires administrative data collected from the NPD. Furthermore, attainment was a secondary outcome of the trial which means the trial was not powered for this analysis.

Mediation analysis

Due to the impact of the Covid pandemic on the Adventure Learning trial, attrition levels were much higher than anticipated in the psychological outcomes (SRL-SRS, SEI & SDQ). As such, the sample size, particularly at longitudinal follow up are lower than ideal. In addition to this, the complex nature of the adventure learning landscape and the understanding we have about how adventure learning can lead to increased attainment means the mediation model suggested is complex. More complex models often benefit from a larger sample size. As such, the SEM analysis is being viewed as exploratory in nature and no causal inferences will be made. Furthermore, prior to conducting mediation analysis, correlational analysis is being conducted to avoid overspecification.

It is possible that unmeasured confounders will be impacting the relationships between variables within the models suggested. As these are unmeasured, we cannot account for them in the specified models.

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), pp. 427-445.

Ballen, C. J., & Salehi, S. (2021). Mediation Analysis in Discipline-Based Education Research Using Structural Equation Modeling: Beyond "What Works" to Understand How It Works, and for Whom. Journal of microbiology & biology education, 22(2), e00108-21. https://doi.org/10.1128/jmbe.00108-21

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

Goodman, R. (2001). 'Psychometric properties of the strengths and difficulties questionnaire', *Journal of the American Academy of Child & Adolescent Psychiatry*, *40*(11), pp. 1337-1345.

Hayes, A. F., & Scharkow, M. (2013). The Relative Trustworthiness of Inferential Tests of the Indirect Effect in Statistical Mediation Analysis: Does Method Really Matter? Psychological Science, 24(10), 1918–1927. https://doi.org/10.1177/0956797613480187

Hooper D, Coughlan J, Mullen M. (2008). Structural equation modelling: guidelines for determining model fit. Electron J Bus Res Methods 6:53–60. 19.

Saris, W, E., Satorra, A., Van der Veld, W,M. (2009) Testing structural equation models or detection of misspecifications? Struct Equ Model 16:561–582. https://doi.org/10.1080/10705510903203433.

Toering, T., et al. (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), pp. 24-38