Transforming Access and Student Outcomes in Higher Education

Final analysis report
Learner analytics
Randomised Controlled Trial
(Sheffield Hallam University)

December 2023

The full protocol for this study is on the TASO website here.

The study was pre-registered on OSF Registries: https://osf.io/y7ku8.

1. Summary

Background: The Behavioural Insights Team (BIT) was commissioned by the Centre for Transforming Access and Student Outcomes in Higher Education (TASO) to act as an independent evaluator of two randomised controlled trials. Both trials were designed to assess the impact of learning analytics interventions. This report corresponds to the trial delivered at Sheffield Hallam University (SHU).

Aims: To evaluate whether a preventative intervention targeted at students identified as being 'at-risk' via SHU's learning analytics programme increases student engagement.

Intervention: Student Support Advisers (SSAs) from a central team proactively monitored engagement at two pre-agreed census points (week 5 and 8 of the autumn term) to identify students who have poor engagement with their course.

- In the intervention 1 group students who generated a red flag (indicating low engagement) in week 4 and/or week 7 received an email detailing support resources available to them plus a text message (SMS) informing them that they will receive a default phone call from a central support team. An SSA then attempted to call all students.
- In the intervention 2 group students who generated a red flag (indicating low engagement) in week 4 and/or week 7 received an email detailing support resources available to them. No telephone calls were made.

Design: This study was a two-arm, parallel group randomised controlled trial, testing for superiority of the intervention 1 condition over the intervention 2 condition.

Outcome measures: There was one primary outcome, the proportion of Red RAG engagement scores at week 9 (defined in section 3.3).

Analyses: A combination of logistic and ordinary least squares (OLS) regressions was used, as appropriate, to estimate effects on the primary and secondary outcomes.

Results: The primary analysis suggests no benefit to students of intervention 1 over intervention 2. All estimated effects are small, and none are statistically significant at the 5% level. The impact table for the results is in Appendix C.

2. Introduction

2.1. Background

This project was a collaboration between the Centre for Transforming Access and Student Outcomes in Higher Education (TASO), Sheffield Hallam University (SHU) and the Behavioural Insights Team (BIT). Between 17 October 2022 - 30 November 2022, Student Support Advisers (SSAs) from a central team at SHU proactively monitored engagement at two pre-agreed census points (week 4 and 7 of the autumn term – week commencing 17th October and 7th November respectively) to identify students who have poor engagement with their course) and delivered two different randomly assigned interventions to these students. BIT conducted an impact evaluation of the effect of the interventions.

BIT was responsible for:

- designing, analysing and reporting for the impact evaluation and
- randomly assigning participants to the intervention 1 or intervention 2 group for the impact evaluation

SHU was responsible for:

- delivering the intervention and
- collecting outcome data

Table 1: Project personnel

Organisation	Name	Role and responsibilities
TASO	Eliza Kozman Rob Summers	Project lead (commissioner) Project manager
The Behavioural Insights Team (BIT)	Anna Bird Patrick Taylor Jess Hunt Pujen Shrestha Tim Hardy Niall Daly Will Cook Laure Bokobza	Policy QA Project lead (evaluation) Project lead (interim) Quantitative analyst Quantitative analyst Quantitative analyst Research QA Research QA

Sheffield Hallam University (SHU)	Carolyn Fearn Helen Parkin Katie Smaylen Felicity Woodhouse	Project lead (intervention / randomised controlled trial delivery) Academic support Intervention delivery Tracking the data Intervention delivery

2.2. Aims

The purpose of this trial was to investigate whether defaulting students to receive a text message (SMS) followed by a phone call from SHU's central team would increase student engagement scores among at-risk students.

Hypothesis: Providing SMS and phone calls by default to students who receive a red flag (indicating low engagement) on SHU's learning analytics system, Data Explorer, will increase engagement scores compared to students who only receive an email.

Research question: What impact does defaulting at-risk students to receive support phone calls have on student engagement scores?

2.3. Intervention

SHU uses a learning analytics programme - Data Explorer - that draws data from Virtual Learning Environment (VLE) activity, assessment, and attendance to track how engaged a student is with their learning.

The programme generates at-risk early warning alerts if a student does not engage virtually or physically with their course for 14 consecutive days during term time.

SHU carried out a previous study on 228 students to pilot the calling regime, which suggested that students require a text message first, to let them know to expect a call, otherwise they are unlikely to answer. The pilot indicated that calls made students more aware of the services available to them, and only 6% thought the call was not useful at all. This randomised controlled trial assessed the impact of this intervention on behavioural (rather than self-reported) outcomes, with a larger sample of students.

In this trial, SHU assessed the impact of providing phone calls by default to students who generate an at-risk alert in week 4 and/or week 7 of the first academic term. The intervention was issued in weeks 5 and 8.[1] In the intervention 1 condition, students received an email with information about support available to them (identical to the intervention 2 email), plus an SMS informing them to expect a support call. A call was then attempted to all students in the group. In the intervention 2 group, students who generated an at-risk alert received the support email but no additional follow up. Across both interventions, students' Academic Advisors were informed that one of their students had generated an at-risk alert, although this was not expected to trigger additional support.

SHU embedded a flagging system within their LA platform that tagged a student as either in intervention 1 or intervention 2 and triggered the appropriate course of action. Staff at the call centre recorded whether or not a phone call took place and shared this data with BIT.

3. Methods

3.1. Design

This study was a two-arm, parallel group randomised controlled trial, testing for superiority of the intervention 1 condition over the intervention 2 condition. Eligible students were randomly assigned to either the intervention 1 group or the intervention 2 group (individual level randomisation).

The monitoring period for outcome data was between 17 October 2022 - 16 December 2022. The planned intervention periods were in week 5 of Term 1 between 24 October 2022 - 30 October 2022, and week 8 in Term 1 between 14 November 2022 - 20 November 2022.

Figure 1 gives an overview of the study flow and timeline up to the point of final data collection. Randomisation was conducted at the level of the student, and so was the analysis.[2]

We considered the risk of spillovers to be low. Given that intervention 1 takes the form of an email sent directly to students, followed by an SMS and a default individualised phone call directed at the treated student's personal phone number, it is unlikely that

---

[1] SHU's plan was to deliver the interventions in week 5 and week 8 of the academic year. However, the intervention based on the week 4 census was delivered on 19 October 2022 (part way through week 4). The intervention based on the week 7 census was delivered on 23 and 24 November 2022 (part way through week 9). This may have resulted in an underestimate of the effect of the intervention on the short-term engagement outcome as the full week 4 census was not complete by the time the first round of the intervention was delivered. However, the estimate on medium-term engagement will not have been affected.

[2] 45 students were identified as being on placement - 44 of those entered the trial. These students were not expected to engage as much and have been removed from the analysis.

untreated students would have been aware or inadvertently benefit from the intervention.

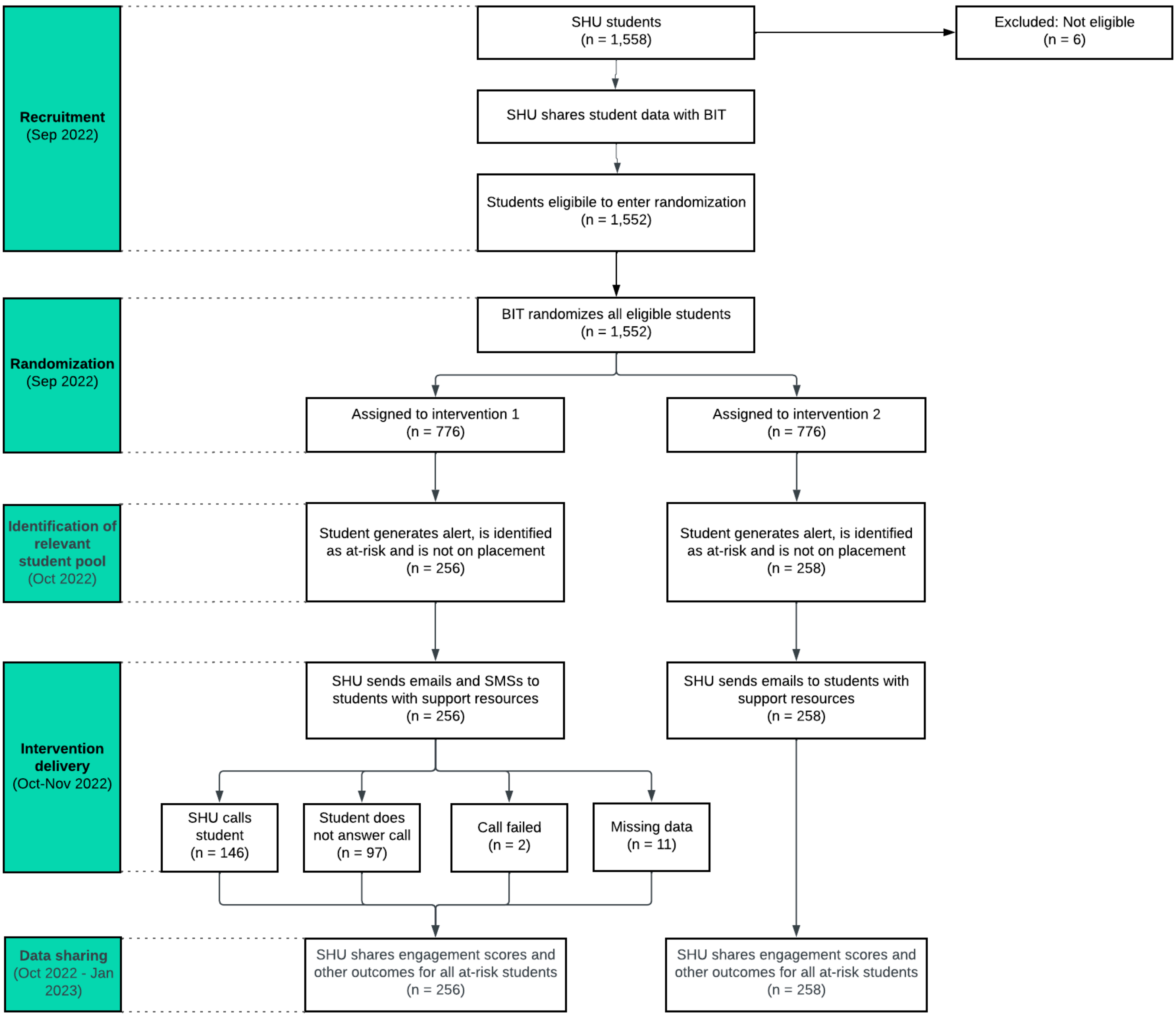

Figure 1: Study flow diagram

3.2. Randomisation

Introduction

The primary practical constraint imposed by SHU on randomisation was that we were not able to randomise the pool of at-risk students directly. Instead, we randomised the whole student population across five courses that agreed to participate in the study, and then included in the analytical sample *only* those who were flagged as at-risk. This introduced the risk that the analytical sample would be unbalanced. To best ensure balance, we conducted a stratified randomisation based on factors that were correlated with being *at risk*, based on conversations with SHU, our own priors from similar research, and the data fields available to us. We determined that in this case, these factors were: year of study; minimum entry qualification (A-levels and equivalent level 3 qualifications, or other qualifications); and ethnicity.

Blinding

Although participants were aware of the at-risk communication they have been exposed to prior to outcome data collection, we do not expect that any were aware that they were in a trial where different participants are exposed to different conditions. This is because students were blind to the randomisation allocation and did not receive any communications about the trial. It should be noted that the intervention delivery staff were not blind to randomisation as they needed to know to which group a participant was allocated in order to deliver the support call to intervention 1.

Allocation mechanism

Randomisation was conducted by BIT. Participants were allocated to a trial arm using a stratified randomisation at the individual-level across the population of undergraduate students in three selected departments, across five courses. They were identified by SHU using a unique student identifier. Analysis was conducted amongst those who were identified as at-risk. We stratified participants on year of study; minimum entry qualification and ethnicity.

Randomisation procedure

SHU provided BIT with a series of spreadsheets containing a list of all students eligible for randomisation. The variables used for randomisation were as follows.[3]

- Participant ID
- Year of study
- Minimum entry qualification
- Ethnicity

[3] An error meant that gender, age and course/department were included as variables which would be used for randomisation in the trial protocol.

This data was shared with BIT through a secure transfer. BIT used the analysis software R to conduct the randomisation, then returned assignment lists to SHU for implementation.

3.3. Outcome measures

The outcomes of interest are described in Table 2. They are broken down into three categories: primary, secondary, and exploratory, defined as follows:

- **Primary outcome:** The main change that the intervention is trying to make.
- **Secondary outcomes:** The other changes that the intervention is trying to make, that are also considered to be valuable ends in themselves.
- **Exploratory outcomes:** Outcomes of interest, but for which we have no strong hypothesis on whether intervention 1 will make a difference.

These definitions are used here to help clarify the intervention's theory, but also to highlight some important analytic choices. The primary outcome was used as the basis for power calculations and the primary/secondary distinction was used to make choices about adjustments for multiple comparisons. The headline findings of the impact evaluation are the estimated effects on the primary and secondary outcomes.

Table 2: Outcome measures

Outcome measure	Data to be collected	Point of collection
PRIMARY: Proportion of Red RAG engagement scores at week 9	Provided by SHU Learning Analytics System	Week 9 Term 1
SECONDARY: Proportion of Red RAG engagement scores at week 12	Provided by SHU Learning Analytics System	Week 12 Term 1
SECONDARY: Whether any additional at-risk flags were generated in Term 1[4]	Provided by SHU Learning Analytics System	Week 12 Term 1
EXPLORATORY: Withdrawal from SHU	Provided by SHU Learning Analytics System	Week 12 Term 1
Notes: Green highlighting indicates primary outcome.		

[4] This outcome has been changed from the Trial Protocol, where it was specified as the number of at-risk flags. Since all students eligible for the analysis have either 1 or 2 at-risk flags, converting this outcome into a binary variable does not lose any information but makes results more intuitive.

The impact evaluation had one primary outcome, the proportion of Red *RAG* engagement scores at week 9. This outcome can be understood as the proportion of Red RAG sub-scores over all Red, Amber, and Green RAG sub-scores at week 9. The outcome can be defined as:

$$RAG_i = \frac{R_i}{R_i + A_i + G_i}$$

where,

- RAG_i is the proportion of RAG sub-scores that are *Red* over all other *Red*, *Amber*, and *Green* RAG sub-scores for participant *i* at week 9

- R_i is the number of sub-scores that are equal to *Red*

- A_i is the number of sub-scores that are equal to *Amber* and

- G_i is the number of sub-scores that are equal to *Green*.

The RAG sub-scores are collected by SHU's Learning Analytics system, Data Explorer, which involves weekly reporting of individual-level data. To generate the RAG sub-scores, each student is evaluated against their module activity in three categories:

- Virtual engagement (logins to the virtual learning environment);
- Physical attendance (tracked via registers in taught timetabled sessions); and
- Assessments (confirmed assessment marks after the semester 1 Boards, end February 2023)[5].

For each module of study a student was given a weekly RAG sub-score of either Red, Amber, Green, or Grey for each engagement category. A RAG sub-score was awarded based on a student's engagement compared to a set of thresholds for a given category. Virtual engagement and physical engagement categories have two engagement thresholds (the number days of no engagement and the percentage engagement less than the average). Whichever RAG score generated from the thresholds is more extreme (Red > Amber > Green > Grey) is the RAG score that is awarded for that category. If there is no activity on the module, then students will receive a grey rating, meaning insufficient data to provide a rating. These thresholds are presented in Table 3. The total number of sub-scores can vary across students. The number of RAG

[5]Assessment RAG sub-scores do not contribute to the RAG overall scores as the monitoring period in this trial is pre-February 2023.

sub-scores will most commonly be 9 or 12 as students will most often be registered on three or four modules.

Table 3: RAG thresholds

	Virtual Engagement		Physical Engagement		Assessment
	Days of no engagement	Percentage engagement less than the average	Days of no engagement	Percentage engagement less than the average	Average mark
Red	**≥ 14 days**	**= 100%**	**≥ 14 days**	**≥ 40%**	**≤ 39**
Amber	**< 14 days and ≥ 7 days**	**< 100% and > 20%**	**< 14 days and ≥ 7 days**	**< 40% and > 20%**	**≤ 40 and ≥ 59**
Green	**< 7 days**	**< 20%**	**< 7 days**	**< 20%**	**≤ 60 and ≥ 100**
Grey	**Insufficient data**	**Insufficient data**	**Insufficient data**	**Insufficient data**	**Insufficient data**

We also analysed two secondary outcome measures: the proportion of Red RAG engagement scores at week 12, and whether a student generated any additional "at risk" flags in Term 1 (they can generate up to two).

3.4. Sample selection

The participant pool was composed of SHU students not on placement in three undergraduate departments[6] who received at least one at-risk flag in Data Explorer over the 2022/23 autumn term (n = 514). Data Explorer identifies students at risk based on their engagement with a range of institutional systems on a daily basis.

The system generated an at-risk early warning alert when a student did not engage with their course for 14 consecutive days during term time, either virtually or physically (or both). A student entered our sample if they generated an alert at week 4 or 7 in the first term.

[6] Courses were selected on the basis that they have large student numbers 600+ and collect individual student engagement data on attendance and module attainment. The undergraduate courses chosen have some of the largest cohort sizes at SHU and all have semester 1 modules, meaning that assessment outcomes from semester 1 could be confirmed at Department Assessment Boards in February 2023. The subject areas/courses chosen have similar delivery patterns in terms of academic time and study patterns (eg. no block study or work-based learning activities).

3.5. Analytical strategy[7]

Primary outcome: Student engagement score at week 9

We used the following model to estimate the effects of the intervention on the primary outcome, using ordinary least squares (OLS) regression. Analysis was conducted on an intention-to-treat basis, including all complete cases.

$$Y_i = \beta_0 + \beta_1 T_i + \beta_2 X_i + \epsilon_i$$

where,

- Y_i is the short-term engagement score (% Red RAG engagement score at week 9)
- T_i is a binary indicator for intervention assignment (1 for intervention 1, 0 for intervention 2)
- X_i is a vector of pre-intervention covariates [gender, ethnicity, age, postcode-level marker of disadvantage (POLAR4 quintiles), fee status (Home vs. EU or International), disability status, department, date intervention was delivered, week 3 % RAG engagement score, year of study, minimum entry qualification[8]] and
- ϵ_i is the heteroskedasticity robust residual error term.

The β_i represent the regression coefficients. β_0 gives the value of the mean of the outcome in the intervention 2 group. β_1 gives the value of the estimated treatment effect. β_2 is the vector of the regression coefficients for the covariates.

Secondary outcome 1: Student engagement score at week 12

We used the following model to estimate the effects of the intervention on the first secondary outcome, using ordinary least squares (OLS) regression. Analysis was conducted on an intention-to-treat basis, including all complete cases.

[7] Due to an error in the published Trial Protocol, we have updated this section. The pre-specified "Secondary outcome 3: Whether a phone call took place" was actually not part of the trial, and the specification for "Exploratory outcome: Withdrawal from university" was omitted by accident.
[8] Here and in the following models, this set of covariates has been updated from the Trial Protocol to include variables that the randomisation was stratified on (year of study, minimum entry qualification). Also, 'department' was originally labelled as 'school ID'.

$$Y_i = \beta_0 + \beta_1 T_i + \beta_2 X_i + \epsilon_i$$

where,

- Y_i is the medium-term engagement score (% Red RAG engagement score at week 12)
- T_i is a binary indicator for intervention assignment (1 for intervention 1 at either/both those points, 0 for intervention 2)
- X_i is a vector of pre-intervention covariates [gender, ethnicity, age, postcode-level marker of disadvantage (POLAR4 quintiles), fee status (Home vs. EU or International), disability status, department, date intervention was delivered, week 3 % RAG engagement score, year of study, minimum entry qualification] and
- ϵ_i is the heteroskedasticity robust residual error term.

The β_i represent the regression coefficients. β_0 gives the value of the mean of the outcome in the intervention 2 group. β_1 gives the value of the estimated treatment effect. β_2 is the vector of the regression coefficients for the covariates.

Secondary outcome 2: Additional at-risk flag generated in Term 1

We used the following model to estimate the effects of the intervention on the second secondary outcome, using a logistic regression. Analysis was conducted on an intention-to-treat basis, including all complete cases.

$$Y_i \sim bernoulli(p_i)\ ;\ logit(p_i) = \beta_0 + \beta_1 T_i + \beta_2 X_i$$

where,

$$logit(p_i) = log\frac{p_i}{1-p_i}$$

and,

- Y_i is a binary indicator of whether a student generated an additional at-risk flag in Term 1 (1 if yes; 0 if not)
- p_i is the probability of Y_i

- T_i is a binary indicator of intervention assignment (1 for intervention 1, 0 for intervention 2)
- X_i is a vector of pre-intervention covariates [gender, ethnicity, age, postcode-level marker of disadvantage (POLAR4 quintiles), fee status (Home vs. EU or International), disability status, department, week 3 % RAG engagement score, year of study, minimum entry qualification][9].

The β_i represent the regression coefficients. β_0 gives the value of the mean of the outcome in the intervention 2 group. β_1 gives the value of the estimated treatment effect. β_2 is the vector of the regression coefficients for the covariates.

Exploratory outcome: Withdrawal from university

We used the following model to estimate the effects of the intervention on the exploratory outcome, using a logistic regression. Analysis was conducted on an intention-to-treat basis, including all complete cases.

$$Y_i \sim bernoulli(p_i)\ ;\ logit(p_i) = \beta_0 + \beta_1 T_i + \beta_2 X_i$$

where,

$$logit(p_i) = log\,\frac{p_i}{1-p_i}$$

and,

- Y_i is a binary indicator of whether a student withdrew from university in Term 1 (1 if yes; 0 if not)
- p_i is the probability of Y_i
- T_i is a binary indicator of intervention assignment (1 for intervention 1, 0 for intervention 2)
- X_i is a vector of pre-intervention covariates [gender, ethnicity, age, postcode-level marker of disadvantage (POLAR4 quintiles), fee status (Home vs. EU or

[9] Note that the date the intervention was delivered is not used as a covariate for this outcome, because it is collinear with the outcome: all students who were first sent an email in the second week cannot have had an additional at-risk flag.

International), disability status, department, date intervention was delivered, week 3 % RAG engagement score, year of study, minimum entry qualification].

The β_i represent regression coefficients. β_0 gives the value of the regression intercept.. β_1 gives the value of the estimated treatment effect as a log-odds ratio. β_2 is the vector of the regression coefficients for the covariates.

Complier Average Causal Effect (CACE)

In addition to the pre-specified analysis above, this report includes a Complier Average Causal Effect (CACE) analysis. One-sided 'non-compliance' occurred in intervention 1 in the sense that some students in the sample did not answer the default phone call. The purpose of the CACE analysis is to estimate the effect of actually receiving a phone call for participants in either intervention group who would comply with the random assignment - that is, those who would actually answer the phone call if assigned to intervention 1.

We can estimate the CACE using an instrumental variables approach. The instrumental variable is assignment to intervention 1, which is assumed to influence actual participation in the relevant intervention (i.e. answering the phone call), but not the outcome variable.

Two key assumptions need to hold for this approach:

1. Instrument relevance: Being assigned to intervention 1 increases participation in intervention 1. In this instance, individuals may only answer a support phone call if they are assigned to intervention 1. This is a safe assumption as BIT defined assignment and SHU had control over the support phone calls.
2. Instrument exogeneity: Assignment does not, in itself, have an effect on the outcome of interest - instead, any effect would be achieved through participation in the intervention. This assumption also holds because assignment was random.

The CACE estimations used a two-stage least squares (2SLS) approach:

$$T_i = \alpha_0 + \alpha_1 Z_i + \alpha_2 X_i + \eta_i \quad (1)$$

$$Y_i = \beta_0 + \beta_1 \hat{T}_i + \beta_2 X_i + \epsilon_i \quad (2)$$

where:

- Z_i is a binary indicator for intervention assignment (1 if the individual is assigned to intervention 1 and 0 if they are assigned to intervention 2)
- T_i is whether a student answers at least one phone call
- X_i is a vector of pre-intervention covariates [gender, ethnicity, age, postcode-level marker of disadvantage (POLAR4 quintiles), fee status (Home vs. EU or International), disability status, department[10], date intervention was delivered, week 3 % RAG engagement score, year of study, minimum entry qualification]
- η_i is the error term in the first stage
- Y_i is the outcome of interest.
- $\hat{T}_i$ are the predicted levels of T_i from (1) and
- ϵ_i is the error term in the second stage.

Note that the 2SLS equations above are linear, so we use linear models even for the binary outcomes (for which we otherwise use logistic models). We use robust standard errors for all outcomes.

4. Results

4.1. Participant flow

Table 4 presents the proportion of the randomised sample that generated an at-risk flag and entered the analysed sample. The proportion of participants in the randomised sample is generally balanced across both intervention groups, 33.0% in intervention 1 and 33.2% in intervention 2. Figure 1 (presented previously) presents a CONSORT flow diagram of the trial, with an overview of the timings and sample numbers for recruitment, intervention delivery and outcome collection. The sample varies substantially in terms of size as the analysed sample is a subset of the randomised sample that generated an at-risk flag pre-trial. Attrition between identifying the at-risk sample and outcome analysis was very low (1% or less, depending on the outcome).

Table 4: Summary of proportion of participants that generated an at-risk flag

[10] Updated from trial protocol, "department" was originally labelled as "school ID".

		Intervention 1	Intervention 2	Total
Number of students	Randomised	776	776	1,552
	Entered trial as at-risk (and not on placement)[11]	256	258	514
	Proportion of participants that generated an at-risk flag (and were not on placement)	33.0%	33.2%	33.1%

4.2. Description of data

Sample demographics

Table 5 shows the baseline demographic characteristics for each intervention group in the two samples: the randomised sample and the analysed sample. A series of chi-squared tests (see Table A1 in Appendix A) on the demographic characteristics of the randomised sample and analysed sample revealed that there are significant differences between the samples for three of the recorded characteristics; gender, year of study, and department. The analysed sample contains a larger proportion of male students, second year students, Finance, Accounting & Business students, and Management students, suggesting that these participants were more likely to be at-risk of disengagement. Nonetheless, in each case the effect size of the difference, assessed using Cramér's V, is weak or non-existent.[12]

Table 5: Distribution of covariates by trial group (Intervention 1 = automatic phone call; Intervention 2 = email only).

	Randomised sample		Analysed sample	
	Intervention 1 (N = 776)	Intervention 2 (N = 776)	Intervention 1 (N = 256)	Intervention 2 (N = 258)
Gender				
Female	408 (52.6%)	410 (52.8%)	99 (38.7%)	97 (37.6%)

[11] 45 students were identified as being on placement - 44 of those entered the trial. These students were not expected to engage as much and have been removed from the analysis.
[12] Interpretation of Cramérs V is dependent on the number of categories (Cohen, 1988) but in all cases reported here values <0.07 indicate no effect. In general for 2x2 contingency tables, 0.10-0.29 is a small effect, 0.3-0.49 is a medium effect and >0.5 is a large effect.

Male	368 (47.4%)	366 (47.2%)	157 (61.3%)	161 (62.4%)
Ethnicity				
Asian	93 (12.0%)	102 (13.1%)	38 (14.8%)	45 (17.4%)
Black	36 (4.6%)	35 (4.5%)	13 (5.1%)	18 (7.0%)
Other	66 (8.5%)	59 (7.6%)	25 (9.8%)	19 (7.4%)
White	581 (74.9%)	580 (74.7%)	180 (70.3%)	176 (68.2%)
Minimum entry qualification				
A-levels/equivalent	543 (70.0%)	542 (69.8%)	175 (68.4%)	176 (68.2%)
Other qualification	233 (30.0%)	234 (30.2%)	81 (31.9%)	82 (31.8%)
Year of study				
1	260 (33.5%)	260 (33.5%)	58 (22.7%)	67 (26.0%)
2	217 (28.0%)	217 (28.0%)	98 (38.3%)	94 (36.4%)
3	237 (30.5%)	252 (32.5%)	71 (27.7%)	71 (27.5%)
4	51 (6.6%)	42 (5.4%)	25 (9.8%)	23 (8.9%)
5	9 (1.2%)	3 (0.4%)	3 (1.2%)	2 (0.8%)
6	2 (0.3%)	2 (0.3%)	1 (0.4%)	1 (0.4%)
POLAR4 Quintile				
1	171 (22.0%)	183 (23.6%)	58 (22.7%)	55 (21.3%)
2	200 (25.8%)	186 (24.0%)	71 (27.7%)	62 (24.0%)
3	108 (13.9%)	139 (17.9%)	36 (14.1%)	47 (18.2%)
4	119 (15.3%)	86 (11.1%)	31 (12.1%)	29 (11.2%)
5	142 (18.3%)	144 (18.6%)	50 (19.5%)	53 (20.5%)
Missing	36 (4.6%)	38 (4.9%)	10 (3.9%)	12 (4.7%)
Department				

Fin, Acc & Bs	178 (22.9%)	168 (21.6%)	82 (32.0%)	74 (28.7%)
Law & Crim	333 (42.9%)	334 (43.0%)	66 (25.8%)	73 (28.3%)
Management	265 (34.1%)	274 (35.3%)	108 (42.2%)	111 (43.0%)
Mature / Young*				
Mature	70 (9.0%)	72 (9.3%)	24 (9.4%)	26 (10.1%)
Young	706 (91.0%)	704 (90.7%)	232 (90.6%)	232 (89.9%)
Disability status				
Disabled	148 (19.1%)	143 (18.4%)	42 (16.4%)	49 (19.0%)
No Disability / Unknown	628 (80.9%)	633 (81.6%)	214 (83.6%)	209 (81.0%)
Fee Status				
Home	743 (95.7%)	742 (95.6%)	248 (96.9%)	249 (96.5%)
Other	33 (4.3%)	34 (4.4%)	8 (3.1%)	9 (3.5%)
Notes: Totals do not add up to 100% due to rounding. * A participant is considered 'mature' if they are over the age of 21 on entry to university.				

Balance checks

Table 6 presents balance checks on the analysed sample. To assess balance, we calculate the differences in mean scores between the two groups for each covariate.[13] Rather than reporting simple differences in means for each covariate, normalised differences are presented to aid comparison between covariates that have different units, and to facilitate comparisons across studies.

We pre-specified balance checks for the following covariates and baseline characteristics:

- Gender

[13] A common alternative is to report whether differences between groups are statistically significant at a certain level of confidence (often $p < 0.05$ in the social sciences). This approach is not particularly helpful because it only tells us whether the sample is large enough to detect a difference, and leaves open the question as to whether any observed differences – and any associated bias – can be addressed through simple covariate adjustment (the approach taken in the analysis for this study) (Imbens & Rubin 2015, p.311).

- Young/Mature student (a binary indicator of whether a student is aged 21 or over at the start of their undergraduate studies)
- Year of study (the year the student commenced their study)
- Department (the department the student is a member of)
- Entry Qualification (a binary indicator of whether a student has entered university with A-levels or another qualification)

The normalised difference is defined as the difference in means between the two groups, divided by the pooled standard deviation. Normalised differences with a magnitude of 0.1 or less indicate a negligible correlation between the covariate and assignment to the intervention 1 group, which can usually be addressed through covariate adjustment in the regression analysis (Austin 2009, p.1233), as done in this report. According to this benchmark, the analytic sample appears to be well-balanced on all pre-specified covariates

Table 6: Balance checks for the analysed sample

	Intervention 1		Intervention 2		
	Mean	**(S.D.)**	**Mean**	**(S.D.)**	**Normalised difference**
Gender (Male)	0.613	0.488	0.624	0.485	-0.022
Young / Mature Student (Young)	0.906	0.292	0.899	0.302	0.024
Year of student (First year)*	0.227	0.458	0.260	0.439	-0.077
Year of student (Second year)*	0.383	0.487	0.364	0.482	0.038
Year of student (Third year)*	0.277	0.449	0.275	0.447	0.005
Year of student (Fourth year)*	0.098	0.297	0.089	0.286	0.029
Year of student (Fifth year)*	0.012	0.108	0.008	0.088	0.040
Year of student (Sixth year)*	0.004	0.062	0.004	0.062	0.000
Department (Finance, Accounting & Business)	0.320	0.468	0.287	0.453	0.073

Department (Law and Criminology)	0.258	0.438	0.283	0.451	-0.056
Department (Management)	0.422	0.495	0.430	0.496	-0.017
Entry Qualification (A-levels)*	0.684	0.466	0.682	0.467	0.003

Notes: N = 514. All variables are binary indicators, so mean averages represent proportions of the group. The parentheses indicate the category of the covariate which was used as the comparison group in the balance check. Reported means and S.D.s are of the non-missing sample.
* Indicates covariates used in stratified randomisation procedure.

Descriptive statistics for outcomes

Table 7 presents the means and standard deviations for the outcomes, broken down by intervention group. In general, it appears that both intervention 1 and intervention 2 performed similarly across all outcomes. In both interventions the proportion of Red RAG engagement scores at week 9 was identical (0.58). The proportion of Red RAG engagement scores at week 12 was lower than in week 9, and slightly higher in intervention 1 than in intervention 2 (0.47 vs. 0.45).

The mean proportion of students who had an additional at-risk flag generated in Term 1 was similar across both groups, with a proportion of 0.27 in intervention 1 and 0.28 in intervention 2. The mean proportion of withdrawals from university was close to 0 across both interventions, with the raw number of withdrawals in the intervention 1 group being 2 and the raw number of withdrawals in the intervention 2 group being 3. This means that regression results for this outcome are not particularly reliable.

Table 7: Average outcome scores by treatment group

Outcome	Intervention 1	Intervention 2
	Mean (SD)	**Mean (SD)**
Proportion of Red RAG engagement scores at week 9	0.58 (0.26)	0.58 (0.26)
N observations	256	255
Proportion of Red RAG engagement scores at week 12	0.47 (0.35)	0.45 (0.32)
N observations	254	255

	Proportion (SD)	Proportion (SD)
Whether an additional at-risk flag was generated in Term 1	0.27 (0.45)	0.28 (0.45)
N observations	256	258
Withdrawal from University in Term 1	0.008 (0.088)	0.012 (0.107)
N observations	256	258
Notes: The N per arm is smaller in some cases than the total analytic sample recorded in the flow diagram. This is because not all students have values for all outcomes. Specifically, observations are missing for the primary and secondary outcomes if a student withdrew before the end of the 9 and 12 weeks respectively.		

4.3. Outcome of analysis

Main pre-specified analysis

Table 8 presents the estimated average effects of participating in intervention 1 versus intervention 2 on the outcomes of interest, for our main model (which includes only complete cases; the full regression tables are in Appendix B). Effects are also presented as standardised effect sizes to make it easier to compare between outcomes and with other studies.

All four of the estimated effects are small - for the proportion of Red RAG engagement scores at weeks 9 and 12 the direction is positive (i.e. increasing non-engagement at university), whereas for the other outcomes the direction is negative (i.e. reducing non-engagement at university). None of the estimates are significant at the 5% level. While this may partly be due to the size of the sample, we cannot conclude with sufficient certainty that the results represent true intervention effects, as opposed to random noise.

Table 8: Estimated effects for the outcomes of interest for intervention 1 (automatic phone call) relative to intervention 2 (email only).

Outcome	Mean for intervention 2	Estimated effect	Standard error	Standardised effect size	Unadjusted p-value
	Linear regression results				
Proportion of Red RAG engagement scores at week 9 (n_1 = 246, n_2 = 243, N = 489)	0.575	0.003	0.022	0.010	0.904

Proportion of Red RAG engagement scores at week 12 (n_1 = 244, n_2 = 243, N = 487)	0.454	0.013	0.028	0.040	0.626
	Logistic regression results				
Additional at-risk flag generated in Term 1 (n_1 = 246, n_2 = 246, N = 492)	0.280	-0.177	0.259	-0.078	0.495
Withdrawal from University in Term 1 (n_1 = 246, n_2 = 246, N = 492)	0.012	-0.145	1.066	-0.015	0.892

Notes: n_1 and n_2 denote the number of individuals in the analysis sample for that outcome for interventions 1 and 2 respectively; N is the total number of individuals in the analysis sample. Observations are missing for the primary and secondary outcomes if a student withdrew before the end of the 9 and 12 weeks respectively.

The standardised effect for linear regression is presented in Hedges's g and the standardised effect for logistic regression is presented in Cohen's h.

+ $p<0.1$, * $p<0.05$, ** $p<0.01$, *** $p<0.001$

Figures 2 through 5 visualise the effects presented in Table 8. The bar lengths for intervention 1 represent what would have happened in the intervention 2 group if they had received intervention 1. Statistically, that means starting from the descriptive mean in the intervention 2 group for the complete case sample and 'adding in' the intervention 1 effect. The uncertainty around the results are illustrated through the orange error bars which indicate a 95% confidence interval.

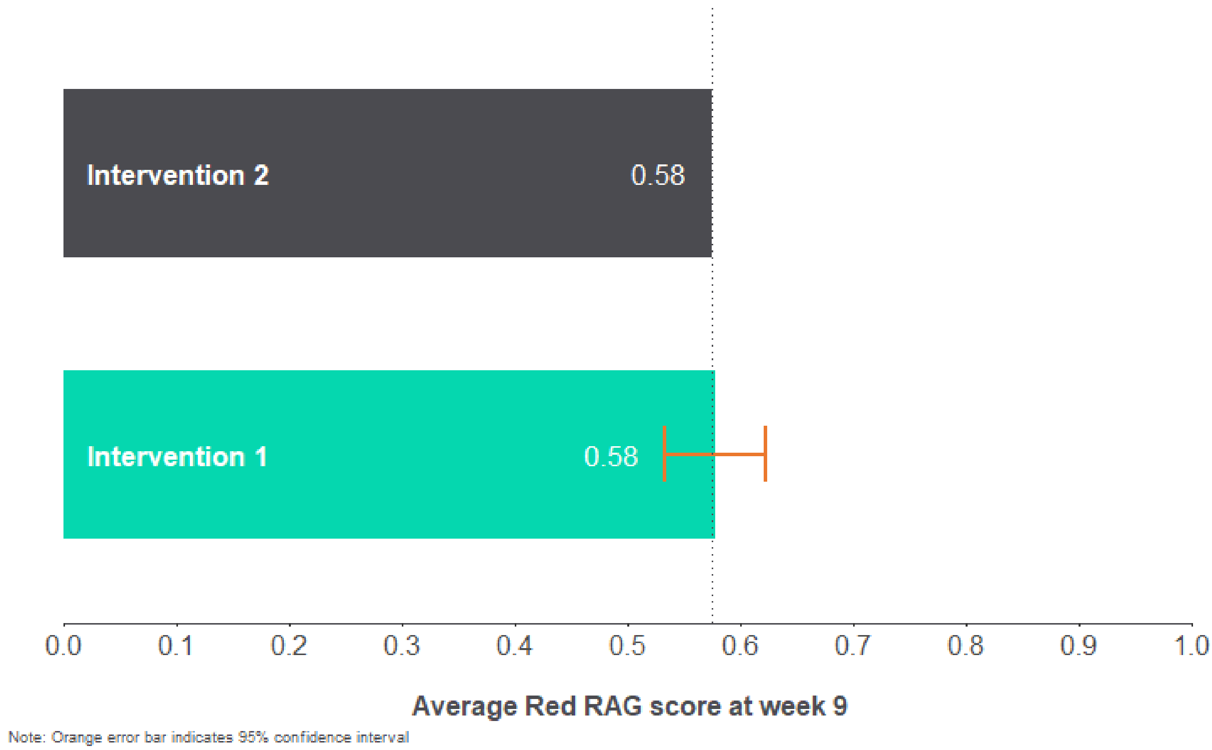

Figure 2: Red RAG score at Week 9 (Intervention 1 = automatic phone call; Intervention 2 = email only).

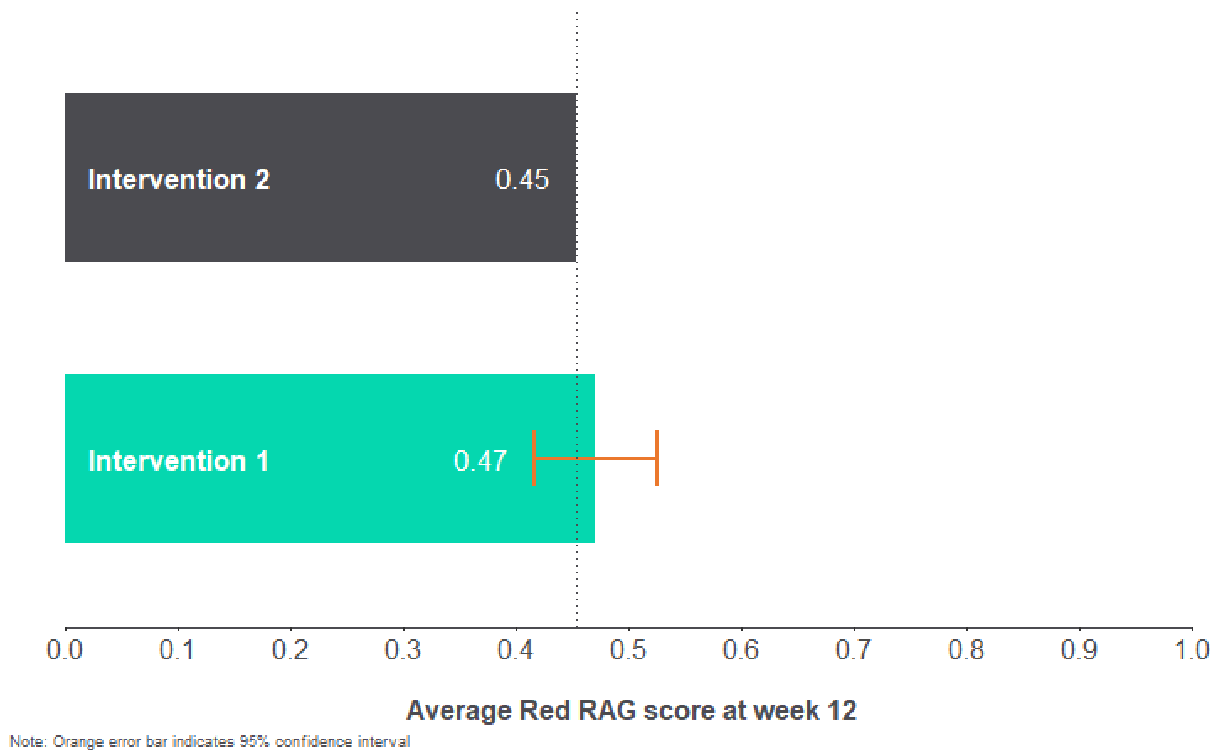

Figure 3: Red RAG score at week 12 (Intervention 1 = automatic phone call; Intervention 2 = email only).

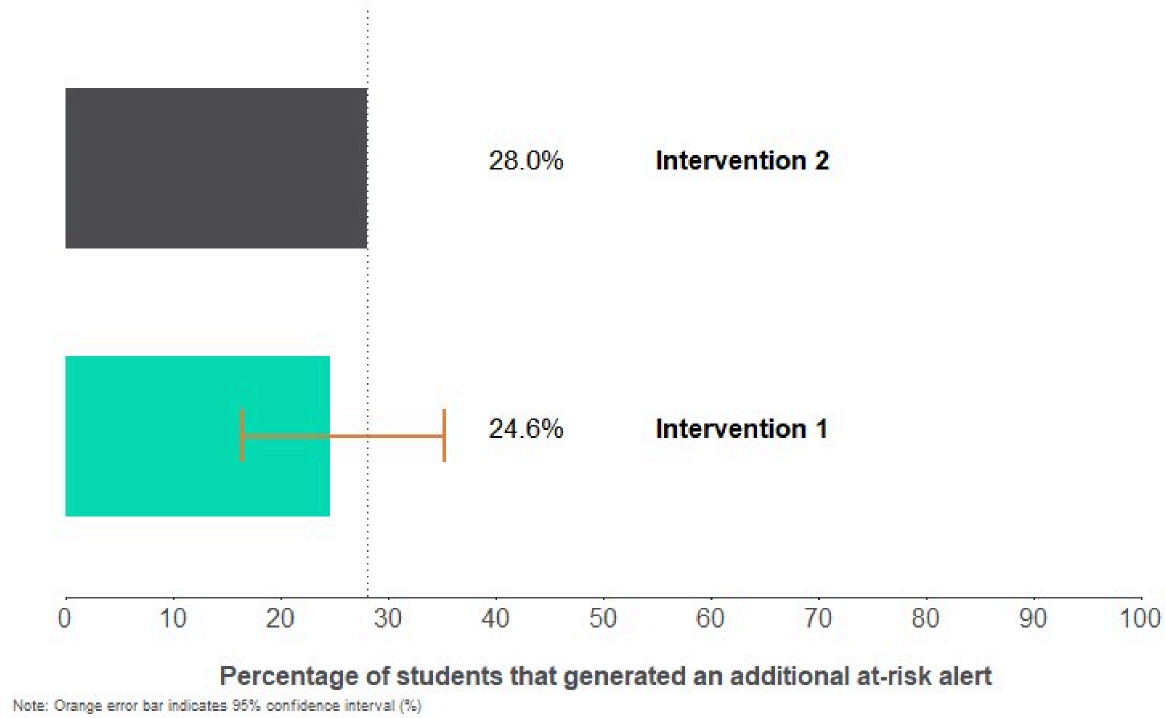

Figure 4: Percentage of students that generated an additional at-risk alert in Term 1 (Intervention 1 = automatic phone call; Intervention 2 = email only).

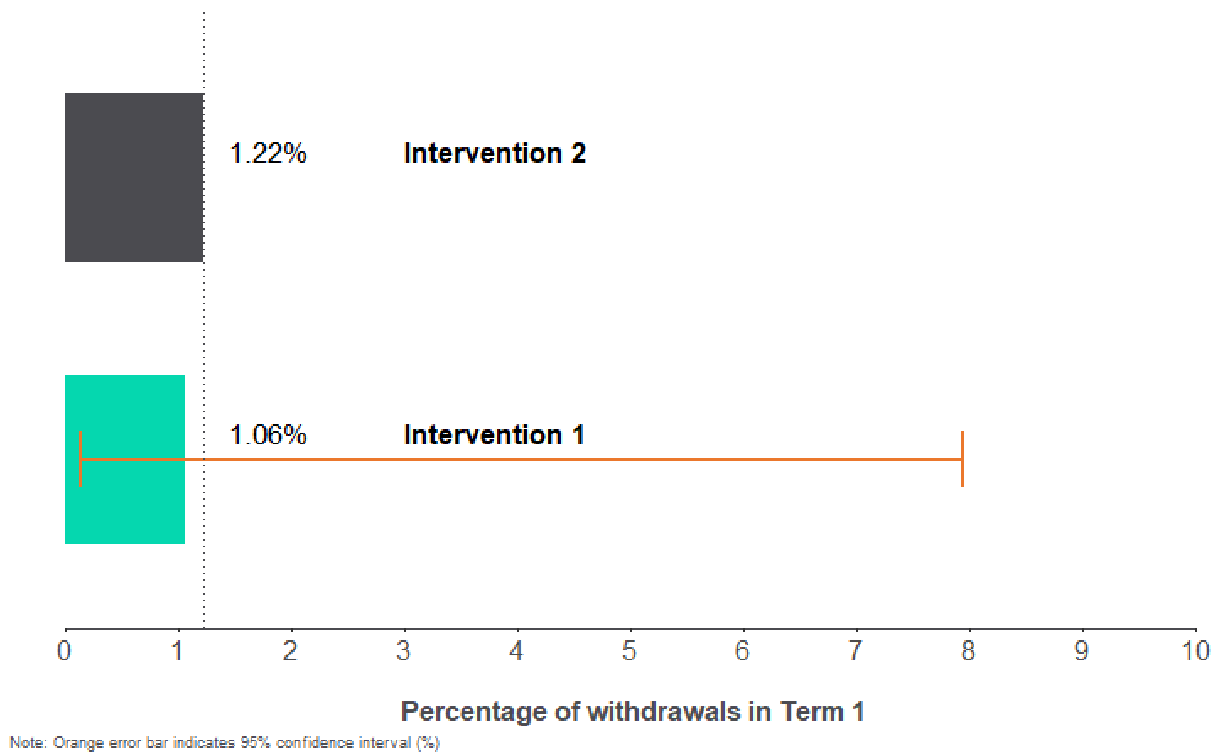

Figure 5: Percentage of withdrawals in Term 1 (Intervention 1 = automatic phone call; Intervention 2 = email only).

Robustness checks

We have run the following robustness checks and find that the results from the pre-specified analysis are broadly robust to these different model specifications.

Missing data

As pre-specified, we have checked whether these results are sensitive to missing data. First, we created a new variable to indicate missingness and used this to re-estimate the effects (Model 2). Second, we re-ran all analyses without covariates (still excluding observations with missing values for any covariate) to obtain the unadjusted estimates (Model 3). Both of these models produce results that are fairly close to those of the primary analysis. Model 2 replicates the direction and (roughly) the magnitude of the effect on intervention 1 in comparison to intervention 2 with respect to the primary outcome. The estimated effect on the primary outcome is reversed in Model 3 but it is still very close to 0, and not close to being significant. There are no major differences in estimates for the other outcomes either.

CACE analysis

We have also conducted a CACE analysis (Model 4) to estimate the impact of receiving a phone call. This estimates the effect of intervention 1 relative to intervention 2 for compliers - that is, those who would actually answer the default phone call. For all outcomes, the two-stage least-squares estimation method is used, which uses linear models. We do not find that receiving a phone call as part of intervention 1 had a significant effect on any of our outcomes of interest.

Table 9 presents the estimated effects from the pre-specified models for each outcome alongside the effects from the alternative models.

Table 9: Estimated effects with different model specifications for intervention 1 (automatic phone call) relative to intervention 2 (email only).

		Estimated effects (SE)			
Outcome		**Model 1**	**Model 2**	**Model 3**	**Model 4**
Proportion of Red RAG engagement scores at week 9	Mean for intervention 2	0.575	0.577	0.575	0.575
	Estimated effect (SE)	0.003 (0.022)	0.006 (0.022)	-0.000 (0.023)	0.005 (0.039)
	N observations	489	511	489	489
Proportion of Red RAG	Mean for intervention 2	0.454	0.452	0.454	0.454

engagement scores at week 12	Estimated effect (SE)	0.013 (0.028)	0.021 (0.027)	0.013 (0.030)	0.024 (0.049)
	N observations	487	509	487	487
Additional at-risk flag generated in Term 1	Mean for intervention 2	0.280	0.279	0.280	0.280
	Estimated effect (SE)	-0.177 (0.259)	-0.123 (0.256)	-0.103 (0.203)	-0.035 (0.059)
	N observations	492	514	492	492
Withdrawal from University in Term 1	Mean for intervention 2	0.012	0.012	0.012	0.012
	Estimated effect (SE)	-0.145 (1.066)	-0.145 (1.066)	-0.410 (0.917)	-0.007 (0.018)
	N observations	492	514	492	492

Notes:

Model 1 = with pre-specified covariates, includes complete cases only (linear model for outcomes related to the proportion of Red RAG scores, logistic otherwise)

Model 2 = with missing covariate data replaced with missingness indicator (linear model for outcomes related to the proportion of Red RAG scores, logistic otherwise)

Model 3 = with no covariates, same sample as model 1 (linear model for outcomes related to the proportion of Red RAG scores, logistic otherwise)

Model 4 = CACE analysis with pre-specified covariates, includes complete cases only (linear models with robust standard errors for all outcomes)

+ p<0.1, * p<0.05, ** p<0.01, *** p<0.001

5. Discussion

Interpretation

The primary intention-to-treat analysis suggests no benefit to students of intervention 1 over intervention 2. All estimated effects from the primary analysis are small, and none are statistically significant at the 5% level. It should be noted that the phone call aspect of the intervention is in the pilot stage of development at SHU. The initial round of calls at week 5, were delivered by a recently trained caller thus potentially leaving space for improvement to the standard of the support calls. The maturation of this aspect of the intervention may lead to a greater impact in the future.

Generalisability

We can think about generalisability in relation to this trial in three ways: i. the extent to which the results might be realised by other universities; ii. the extent to which the results might be realised in different populations; and iii. the extent to which the results might be realised over different time periods in the academic term.

The first two types of generalisation are likely inter-related given that there are a variety of Higher Education Providers in the UK each with their own context, such as the demographics and prior attainment of the student population, and the range and types of courses offered. Universities themselves may therefore share similar aims and approaches to using Learner Analytics, and deliver support with a similar level of quality. Nevertheless, universities can apply their learning analytic tools in various ways. Some universities may focus on those who have no engagement over a specific period of time, others may target those who have limited but still some engagement. The extent to which the effects found in SHU may generalise to other universities will therefore depend in part on the similarity of their student population and the way they use their learning analytics system to identify at-risk students.

On the second type of generalisation, we know that the analysed sample is not wholly representative of the wider student population at SHU. Participants in this trial were from Finance, Accounting & Business, Law & Criminology, and Management departments, which make up only a small proportion of the total departments at SHU and other universities. These departments may be different in terms of contact hours and education delivery methods, which could affect the intensity and regularity with which students engage. Beyond this, student characteristics within these departments may also be different. Therefore the extent to which these departments see a similar effect would partly depend on the extent to which their programme delivery and their cohort of students matched the characteristics of those in this study.

On the third type of generalisation, we know that engagement differs throughout the academic year, as highlighted by the different engagement scores we observed in week 9 and week 12. The demands on students vary across the year, from reading weeks to exam periods. Effects may therefore be different at different times of the year.

Trial limitations

There are three limitations that complicate the interpretation of the results of this study, and one caveat about the nature of who the intervention is attempting to affect.

The first limitation is the design of the Red RAG engagement outcome variables: The aggregation of RAG scores into the composite Red RAG engagement outcome restricts our ability to meaningfully interpret the results.

The composite Red RAG engagement scores outcomes were designed to capture negative engagement performance by students, through the proportion of ‘Red’ RAG scores that students receive. The same composite outcome can therefore reflect a variety of individual situations. For example, a student with 50% Red RAG scores and 50% Green RAG scores would have the same value on the proportion of Red RAG engagement score – the main outcome in the trial – as a student with 50% Red and 50% Amber. However, it is clear that the former student was more engaged than the latter. We would suggest developing outcome measures that more accurately capture positive engagement in future Learner Analytics trials. For example, building on the existing RAG scores as a measure of engagement, each RAG colour could be assigned a numeric weight. Averaging the scores across an individual would generate a weighted engagement score that more accurately reflects the student's degree of engagement.

The second limitation is that the data routinely recorded in the Learning Analytics system for this trial was limited to virtual engagement, physical attendance, and assessments. While these measures make up a large part of potential touchpoints between a student and their university, there are other touchpoints that are overlooked, such as meetings with personal tutors, student-led learning groups, and non-academic extracurricular activities. Acknowledging that engagement at university can mean a range of things, further work exploring the effect of Learner Analytics' on student engagement would benefit from collecting more diverse metrics of engagement. This will produce a more holistic view of engagement at university, and allow future trials to better pinpoint which type of engagement Learner Analytics are best suited to affect, and by how much.

A consequence of these first two limitations is that the design of the index makes it difficult to express and quantify the effect of the intervention in meaningful terms, such as an x% decrease in the risk of dropout, or an increase of y% in attendance. Ensuring that the way in which the main engagement outcome is constructed allows such an interpretation would enhance the impact of the evidence generated by the trial.

Additionally, this trial is limited by the target population of the intervention. As the outcomes of interest focused on students defined as at-risk and on measuring negative engagement, the results only allow us to infer the effect on engagement of defaulting at-risk students to the support phone call *versus* not defaulting them. However, Learner Analytics also has the potential to be used more broadly to help those students whose engagement does not place them at-risk according to the definitions used in this context, but who may still benefit greatly from additional support.

Lastly, a limitation of this trial is the variation in time from at-risk flag generation to intervention delivery across participants. For example, a student who generates an at-risk flag at the start of week 4 would get the intervention at the same time as a student generating an at-risk flag at the end of week 4. This may result in the optimal intervention period, where the intervention would have the greatest effect on engagement, being missed for some participants.

Bibliography

Austin, P.C., 2009. Using the standardized difference to compare the prevalence of a binary variable between two groups in observational research. Communications in Statistics-Simulation and Computation, 38(6), pp.1228-1234.

Cohen, J. 1988. Statistical Power Analysis for the Behavioral Sciences (2nd ed.). Lawrence Erlbaum Associates.

Imbens, G.W. and Rubin, D.B., 2015. Causal inference in statistics, social, and biomedical sciences. Cambridge University Press.

Appendix A: Table of chi-squared tests on the covariates

Table A1:Chi-squared tests between the full sample and the analysed sample for each of the covariates.

	χ^2	Degrees of Freedom	P-value	Cramérs V[#]
Gender	32.235	1	<0.001	0.125
Ethnicity	7.082	3	0.156	0.059
Minimum entry qualification	0.406	1	0.786	0.014
Year of study	30.329	5	<0.001	0.121
POLAR4 Quintile	1.724	5	0.886	0.029
Department	41.974	2	<0.001	0.143
Age	0.092	1	0.857	0.007
Disability status	0.215	1	0.826	0.010
Fee Status	0.767	1	0.686	0.019

p-values are adjusted for multiple comparisons using the Benjamini-Hochberg correction.

[#]Interpretation of Cramérs V is dependent on the number of categories (Cohen, 1988) but in all cases reported here values <0.07 indicate no effect and values <0.15 indicate a small effect. In general for 2x2 contingency tables, 0.1-0.29 is a small effect, 0.3-0.49 is a medium effect and >0.5 is a large effect.

Appendix B: Regression table for pre-specified models (model 1)

Table B1: Full table of regression coefficients for the pre-specified models for short-term engagement, medium-term engagement, whether an additional at-risk flag was generated in Term 1, and withdrawal from university.

	Proportion of Red RAG engagement scores at week 9	Proportion of Red RAG engagement scores at week 12	Whether an additional at-risk flag was generated in Term 1	Withdrawal from University in Term 1
(Intercept)	0.499	0.714	-2.042	-40.200
	s.e. = 0.076	s.e. = 0.086	s.e. = 0.721	s.e. = 7096.719

	p = <0.001	p = <0.001	p = 0.005	p = 0.995
Allocation Ref: Intervention 2				
Intervention 1	0.003	0.013	-0.177	-0.145
	s.e. = 0.022	s.e. = 0.028	s.e. = 0.259	s.e. = 1.066
	p = 0.904	p = 0.626	p = 0.495	p = 0.892
Gender Ref: Female				
Male	0.057	0.043	0.533	-1.108
	s.e. = 0.025	s.e. = 0.031	s.e. = 0.302	s.e. = 1.211
	p = 0.022	p = 0.166	p = 0.078	p = 0.360
Ethnic group Ref: White				
Black	-0.007	-0.037	-0.679	-16.209
	s.e. = 0.055	s.e. = 0.059	s.e. = 0.630	s.e. = 7082.149
	p = 0.900	p = 0.535	p = 0.281	p = 0.998
Asian	0.064	0.063	0.803	1.672
	s.e. = 0.036	s.e. = 0.041	s.e. = 0.359	s.e. = 1.237
	p = 0.074	p = 0.124	p = 0.026	p = 0.177
Other	0.030	0.044	0.380	-16.520
	s.e. = 0.042	s.e. = 0.049	s.e. = 0.459	s.e. = 6369.907
	p = 0.475	p = 0.369	p = 0.407	p = 0.998
Polar4 quintiles Ref: Quintile 1				
Quintile 2	-0.012	-0.010	-0.510	17.556
	s.e. = 0.032	s.e. = 0.041	s.e. = 0.373	s.e. = 3901.802
	p = 0.722	p = 0.798	p = 0.172	p = 0.996
Quintile 3	-0.045	-0.081	-0.139	-0.058

	s.e. = 0.038	s.e. = 0.046	s.e. = 0.444	s.e. = 5953.030
	p = 0.235	p = 0.076	p = 0.755	p = 1.000
Quintile 4	-0.034	0.001	-0.050	19.609
	s.e. = 0.043	s.e. = 0.047	s.e. = 0.446	s.e. = 3901.801
	p = 0.430	p = 0.982	p = 0.911	p = 0.996
Quintile 5	-0.023	0.055	0.405	18.992
	s.e. = 0.038	s.e. = 0.047	s.e. = 0.407	s.e. = 3901.802
	p = 0.552	p = 0.242	p = 0.319	p = 0.996
Mature or young student Ref: Mature				
Young	0.019	-0.068	0.353	17.694
	s.e. = 0.051	s.e. = 0.056	s.e. = 0.488	s.e. = 5927.846
	p = 0.714	p = 0.225	p = 0.469	p = 0.998
Fee status Ref: Home student				
Other	0.156	0.243	-2.018	-0.866
	s.e. = 0.139	s.e. = 0.138	s.e. = 1.527	s.e. = 21845.898
	p = 0.263	p = 0.079	p = 0.186	p = 1.000
Disability status Ref: Disabled				
No Disability/Unknown	-0.069	-0.154	-0.288	0.347
	s.e. = 0.033	s.e. = 0.038	s.e. = 0.341	s.e. = 1.267
	p = 0.037	p = <0.001	p = 0.399	p = 0.784
Date of email Ref: 19/10/2022				
23/11/2022 or 24/11/2022	0.134	-0.023	N/A (not in regression)	-1.070
	s.e. = 0.033	s.e. = 0.038		s.e. = 1.387

	p = <0.001	p = 0.542		p = 0.440
Department Ref: FIN, ACC & BS				
Law & Crim	0.032	-0.094	-1.235	0.189
	s.e. = 0.035	s.e. = 0.043	s.e. = 0.392	s.e. = 1.311
	p = 0.367	p = 0.030	p = 0.002	p = 0.885
Management	-0.089	-0.151	-0.990	-0.855
	s.e. = 0.027	s.e. = 0.031	s.e. = 0.311	s.e. = 1.324
	p = 0.001	p = <0.001	p = 0.001	p = 0.518
Baseline outcome (engagement at week 3)	0.207	0.231	4.418	1.924
	s.e. = 0.057	s.e. = 0.064	s.e. = 0.456	s.e. = 2.201
	p = <0.001	p = <0.001	p = <0.001	p = 0.382
Minimum entry qualification **Ref: A-levels/equivalent**				
Other qualification	-0.025	0.036	-0.208	0.392
	s.e. = 0.026	s.e. = 0.032	s.e. = 0.301	s.e. = 1.154
	p = 0.333	p = 0.260	p = 0.490	p = 0.734
Year of study **Ref: Year 1**				
Year 2	-0.020	-0.058	-0.416	-0.691
	s.e. = 0.027	s.e. = 0.036	s.e. = 0.338	s.e. = 1.393
	p = 0.458	p = 0.110	p = 0.219	p = 0.620
Year 3	-0.028	-0.146	-0.119	-1.055
	s.e. = 0.031	s.e. = 0.040	s.e. = 0.356	s.e. = 1.601
	p = 0.355	p = <0.001	p = 0.738	p = 0.510
Year 4	-0.075	-0.227	-1.237	0.155

	s.e. = 0.049	s.e. = 0.059	s.e. = 0.549	s.e. = 1.597
	p = 0.126	p = <0.001	p = 0.024	p = 0.923
Year 5	-0.153	-0.081	-0.432	-19.664
	s.e. = 0.148	s.e. = 0.183	s.e. = 1.227	s.e. = 17751.746
	p = 0.303	p = 0.657	p = 0.725	p = 0.999
Year 6	0.394	0.063	13.557	-2.910
	s.e. = 0.069	s.e. = 0.305	s.e. = 574.531	s.e. = 34338.698
	p = <0.001	p = 0.836	p = 0.981	p = 1.000
Num. Obs.	489	487	492	492

Appendix C: Impact table

Outcome	Sample size	P Value	Effect	Estimated 'real world' effect	Evaluation security (1 = not at all secure 5 = very secure)	Type of evidence
What is the outcome measure? (include primary and secondary outcomes)	*How many participants were included in the study relating to this outcome?*	*Report the p-value derived from the statistical tests*	*Report the size of the effect - confidence intervals/Cohen's d / Cohen's h*	*Where possible, please translate the effect size into a tangible example of the size of the effect - e.g., 13 more students apply to HE*	*See evaluation security note[14]*	*Is it Type 1,2 or 3 evidence - according to the [OfS standard of evidence]?*
PRIMARY: Proportion of Red RAG engagement scores at week 9	489	0.904	0.010 (Hedges g)	-	3.7	3
SECONDARY: Proportion of Red RAG engagement scores at week 12	487	0.626	0.040 (Hedges g)	-	3.7	3
SECONDARY: Whether any additional at-risk flags were generated in Term 1	492	0.495	-0.078 (Cohen's h)	-	3.7	3

[14] *Based on the decisions made around the evaluation, you will be able to assess the security of your evaluation – that is, how confident you can be when making claims about the findings. The most robust evaluations with large samples, low attrition levels and no threats to validity will receive the highest score of 5/5.*