TASO Transforming Access and Student Outcomes in Higher Education

Final analysis report
Learning analytics
Randomised controlled trials
(Nottingham Trent 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/ygtdx.

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 Nottingham Trent University (NTU).

Aims: To evaluate whether a preventative intervention targeted at students that generate a no-engagement alert via NTU's learning analytics student dashboard (StREAM) increased student engagement.

Intervention:

- In the intervention 1 group, students who generated a no-engagement alert received up to two phone call attempts from NTU's central support team (business as usual).
- In the intervention 2 group, students who generated a no-engagement alert received an email inviting them to request a phone call.

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 were two primary outcomes (detailed in Section 3.3):

I. Average daily student engagement rating in the 10 day period following their first no-engagement flag (days 1 to 10 of the intervention period) and
II. Average daily student engagement rating in the first four-week period of Term 2.

These outcomes were collected by NTU's learning analytics system, which involves daily reporting on individual-level engagement data.

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 (automatic phone call) over intervention 2 (email). Estimated effects on the primary outcomes and first secondary outcome are either null or narrowly negative, 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), Nottingham Trent University (NTU) and the Behavioural Insights Team (BIT). Between 11 October 2022 - 14 December 2022, NTU proactively monitored individual-level engagement to identify students who had no engagement, 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
- providing the code to NTU so they could randomly assign participants to the intervention 1 group or intervention 2 group for the impact evaluation

NTU was responsible for:

- delivering the intervention
- collecting outcome data

Table 1: Project personnel

Organisation	Name	Role and responsibilities
TASO	Eliza Kozman Rob Summers	Project lead (commissioner) Project manager
BIT	Anna Bird Patrick Taylor Jess Hunt Pujen Shrestha Tim Hardy Will Cook Laure Bokobza	Policy QA Project lead (evaluation) Project lead (interim) Quantitative analyst Quantitative analyst Research QA Research QA
NTU	Eleanor Turpin Emma Hynd George Cox Jonathan Hale	Project manager (intervention / randomised controlled trial delivery) Project support and NTU final decision maker IPE author, qualitative research lead and intervention administrator Intervention implementation lead

	Mike Kerrigan Ed Foster	Project support and advice Project support and advice

2.2. Aims

The purpose of this trial was to investigate whether providing a default coaching-style phone call for all students that generated a no-engagement alert, rather than providing the opportunity to voluntarily opt-in to such calls, increased engagement rating for these students.

Hypothesis: Providing coaching-style phone calls by default to students who received a no-engagement alert on StREAM increases engagement rating compared to students who received the option to request coaching-style phone calls.

Research questions: What impact does providing a support phone call to non-engaged students by default have on student engagement rating?

2.3. Intervention

NTU uses a learning analytics dashboard - StREAM - that tracks how engaged a student is with their learning by drawing data from the following institutional systems:

- Virtual Learning Environment (VLE) logins
- VLE learning rooms
- Attendance monitoring
- Online submissions
- Online resource use
- Building access
- Library loans

The dashboard generates 'no-engagement' alerts if a student does not interact with any of the institutional systems listed above during term time for 10 consecutive days for first year students and 14 consecutive days for subsequent years. Lack of engagement with these institutional systems is strongly associated with non-progression (Foster & Siddle, 2019). In 2020, as part of the institution's response to the first national COVID-19 lockdown, NTU used data from the dashboard to create a contact service. 'No-engagement' alerts are sent to a team of callers who attempt to contact students with the aim of providing them with an initial coaching-style phone call.

In this trial, we assessed the impact of providing coaching-style phone calls by default to students the first time they generate a no-engagement alert. When an alert was

generated, an email was automatically sent to personal tutors through the dashboard, including the information that a coaching-style intervention would be made in two days' time. At this point, personal tutors could opt their students out of receiving the intervention. This occurred in small numbers, and happened in situations where the intervention was not relevant to the student, e.g. the student was on a study break.

The following morning, students in the intervention group 1 received an email informing them to expect a support call within 24 hours, unless they opt-out (delivery by default). In the intervention 2 group, students who generated a no-engagement alert received an email inviting them to request a phone call (voluntary delivery).

NTU added a stage in the process that prepared the contact-list based upon no-engagement alerts to tag a student as either intervention 1 (automatic phone call) or intervention 2 (email). They recorded the outcome of the phone call attempts (e.g. spoken to student, left voice mail etc.) and shared this data with BIT.

3. Methods

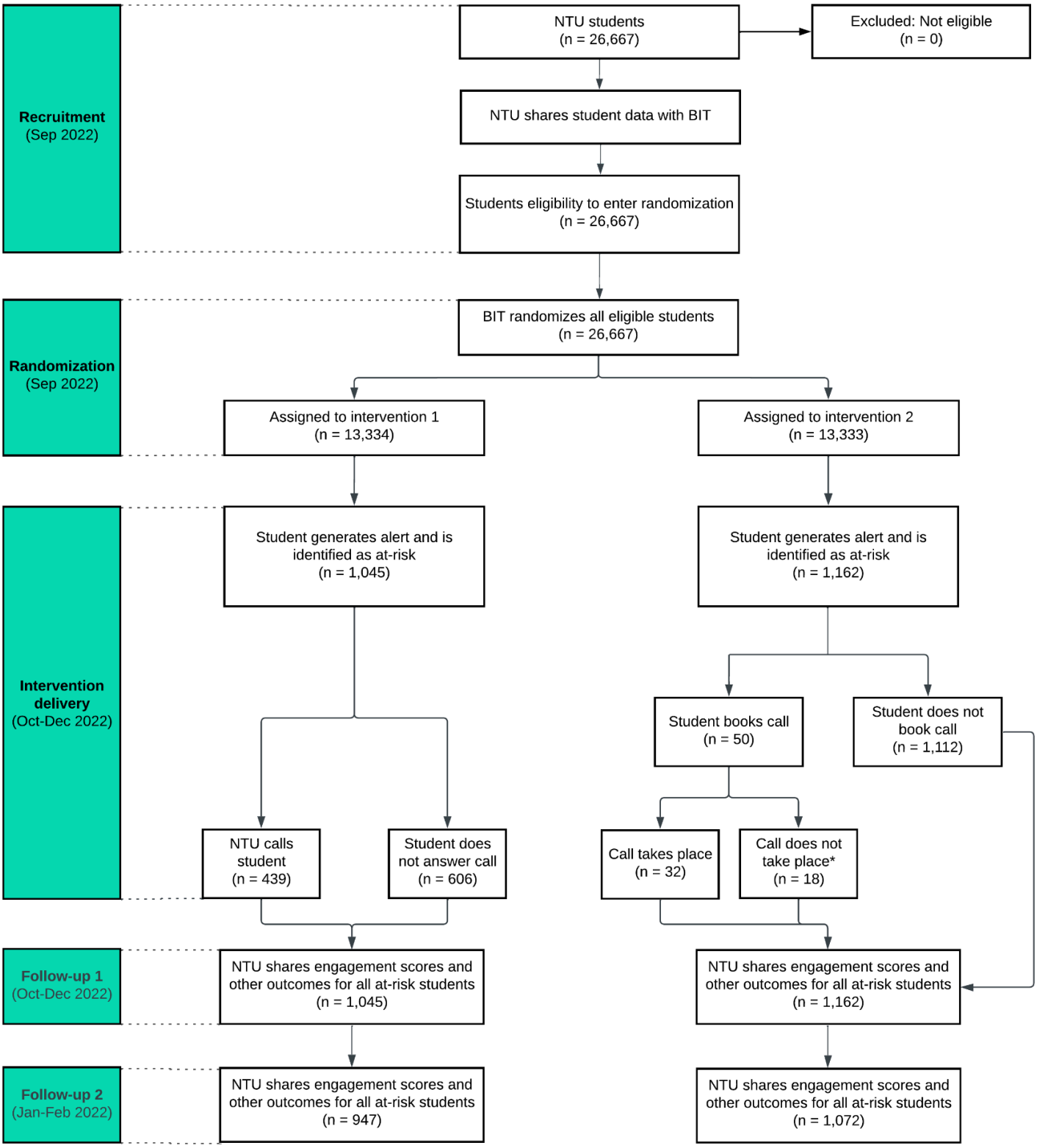
3.1. Design

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 intervention period was between 11 October 2022 and 14 December 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.

We considered the risk of spillovers to be low. Given that intervention 1 takes the form of an email sent directly to students and a default individualised phone call directed at the treated student’s personal phone number, it is unlikely that students in the intervention 2 group would have been aware or inadvertently benefited from the intervention.



*Note: Of the 18 students assigned to Intervention 2 who booked a call which didn't take place, 16 did not answer the call and 2 were not called.

Figure 1: Study flow diagram

3.2. Randomisation

Introduction

Due to operational practicalities, we were not able to randomise the pool of alert-generating students directly. Instead, we randomised the whole eligible student population and then included in the analytical sample *only* those that generated a non-engagement alert. 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 the generation of a no-engagement alert, and with engagement rating more generally, based on conversations with NTU, our own priors from similar research, and the data fields available to us. We determined that in this case, these factors were: ethnicity, year of study, and highest qualification on entry.

Blinding

Although participants were aware of the communication they received 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 which group a participant was allocated in order to deliver the appropriate communication.

Allocation mechanism

Participants were allocated to a trial arm using stratified randomisation at the individual level across the total alert-generating population of undergraduate students attending NTU. They were identified by NTU using a unique student identifier. Analysis was only conducted amongst those who later generated a non-engagement alert. We stratified participants along ethnicity, year of study, and qualification on entry, which are correlated with the generation of non-engagement alerts, in order to ensure balance within the sample of alert-generating students.

Randomisation procedure

The randomisation was stratified on binary versions of three variables:

- Ethnicity (indicator for being white)
- Year of study (indicator for first year)
- Entry qualification (indicator for A-levels).

Due to the tight time frame of this trial and the nature of the data sharing processes, a decision was made that NTU would conduct the randomisation using R code shared by BIT.

3.3. Outcome measures

Introduction

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 (specifically, we adjust for multiple comparisons using the Benjamini-Hochberg procedure within primary and secondary outcomes separately). The headline findings of the impact evaluation are the estimated effects on the primary and secondary outcomes.

Changes to the outcome measures

Revisions were made to the analysis plan of this project as a result of stages of the intervention delivery that were not considered when drawing up the trial protocol. These changes are captured here.

As a condition of ethical approval for the study but missed in the trial protocol, students in the intervention 2 group who generated a second no-engagement alert were automatically called by NTU (i.e., these students received intervention 1 on top of intervention 2). Therefore, the pre-specified average treatment effect compares students in the intervention 1 group as described above with a mixed group of students - some of whom received intervention 2 only as described in the trial protocol, and some of whom received the intervention 2 condition plus the intervention 1 condition. For simplicity and consistency, **we refer to this mixed group as 'intervention 2' in this report**. To estimate a treatment effect that compares the intervention 1 and intervention 2 conditions as they are described in the protocol, we cannot simply drop participants in the group initially allocated to intervention 2 who started to receive automatic phone calls, because this would introduce selection bias.

We have therefore taken the approach of changing the relevant outcome definitions so that measurements are taken before any second no-engagement alerts could have been generated. Since an alert is generated every 10 to 14 days depending on a

student's year of study, we have measured engagement over the 10 days after a first alert was generated for each student for the relevant outcomes. The benefit of this approach is that we can include the whole sample in the analysis for these outcomes and get an unbiased estimate of the treatment effect as defined in the trial protocol - i.e. a comparison of students who received default phone calls in one group, and students who had to request phone calls in the other group. We are able to estimate treatment effects of this type for the first primary outcome (a student's short-term engagement rating) and the exploratory outcome (attendance).

For the second primary outcome (a student's medium-term engagement rating), the period between a first no-engagement alert and the point of outcome measurement is long enough that there are students in the intervention 2 group who received two or more alerts (and therefore some default phone calls) before the outcome was measured. To address this issue, we conduct two analyses:

1. The **pre-specified analysis, with all complete cases**, with the caveat that the estimated treatment effect for this medium-term outcome represents a comparison of different conditions to that of the short-term outcome (i.e. the intervention 2 group includes students who received default phone calls in this case).

2. A **matching approach** where we drop individuals who received 2 or more no-engagement alerts from each intervention group. This estimate is biassed in expectation as those who remain in the analytic sample do not do so randomly. This analysis may also be underpowered.

We have included an additional version of the outcome "student answers the phone call", to account for the fact that there are two possible versions of this outcome dependent on what one considers the relevant sample of reference to be. Specifically, for the intervention 2 group, the proportion of students that answered a phone call can be calculated as *the number of students who answered the call* divided by *the total number of students who booked calls*. Alternatively, this outcome could be calculated as the *number of students who answered the call* divided by *the total number of students in the intervention 2 group.* We have included the second definition of this outcome as exploratory analysis.

Table 2 summarises the changes that have been made to the outcome measures.

Table 2: Changes to pre-specified outcomes

Outcome measure	Data to be collected	Original collection period	New collection period	Reason for the change
PRIMARY: Student's mean short-term engagement rating	Provided by NTU Learning Analytics System	Mid-line (days 7 to 21 of the intervention period)	Days 1-10 of the intervention period	Collection period reduced in length to avoid capturing the effect of 2nd alerts.
PRIMARY: Student's mean medium-term engagement rating	Provided by NTU Learning Analytics System	Endline (first 4 weeks of Term 2)	Not changed	As it is not possible to reduce the length to avoid capturing the effect of 2nd alerts we have decided not to change the collection period.
SECONDARY: Additional no-engagement alert generated in Term 1	Provided by NTU Learning Analytics System	Mid-line (end of Term 1)	Not changed	This outcome did not require revision because the change in intervention allocation does not affect it.
SECONDARY: Student answers phone call (including in the model all students in intervention 1 and those students in intervention 2 who booked a phone call)	Provided by NTU administration	Mid-line (2 weeks following intervention, Term 1)	Not changed	This outcome is measured when students generate their first no-engagement alert, therefore the change in intervention allocation does not affect it.
EXPLORATORY: Student answers phone call (including in the model all alert-generating students)	Provided by NTU administration	Mid-line (2 weeks following intervention, Term 1)	Not changed	This is an additional outcome. This outcome is measured when students generate their first no-engagement alert, therefore the change in intervention allocation does not affect it. However, we have changed the denominator for calculating the proportion to the full intervention 2 group.

Outcome measure	Data to be collected	Original collection period	New collection period	Reason for the change
EXPLORATORY: Attendance	Provided by NTU administration	Mid-line (2 weeks following intervention, Term 1)	Days 1-10 of the intervention period	Collection period reduced in length to negate the effect of 2nd alerts.
Note: an additional exploratory outcome, Student withdraws from NTU, was included in the trial protocol but the data were not available at the point of analysis				

Non-compliance

There was an administrative error during the trial period on 7 November 2022 which led to members of both intervention groups being given the wrong intervention (two-sided non-compliance[1]). NTU has shared the unique identifiers for each of these participants with BIT. BIT first checked the balance of non-compliance, in an attempt to understand whether this issue systematically affected one arm over the other. The timing of the non-compliance means that it did not affect all the students used to analyse short-term outcomes (student's short-term engagement rating and attendance). We have therefore conducted two balance checks, one for compliance among the short-term outcomes and one for compliance among the other outcomes.

We find that non-compliance occurs in both arms. 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. For both short-term and longer-term outcomes the magnitude of the normalised difference is slightly above 0.1, which means that there was more non-compliance in intervention 2 than intervention 1. To account for non-compliance in the analysis, we have included an additional robustness check where we analyse only the compliant sample.

Table 3: Balance check on compliance. Proportion (and standard deviation (SD)) of students in each intervention group that were non-compliant (Intervention 1 = automatic phone call; Intervention 2 = email only).

	Intervention 1 mean	Intervention 1 SD	Intervention 2 mean	Intervention 2 SD	Normalised difference
Non-compliance	0.027	0.162	0.058	0.233	-0.154

[1] Non-compliance affected a total of n = 170 students, where n = 65 were affected in the intervention 1 group and n = 105 were affected in the intervention 2 group.

for short-term outcomes					
Non-compliance for all other outcomes	0.062	0.242	0.090	0.287	-0.106

Outcomes

There are two primary outcomes for our impact evaluation:

I. Average daily student engagement rating in the 10 day period, following their first no-engagement flag (days 1 to 10 of the intervention period); and
II. Average daily student engagement rating in the first four-week period of Term 2.

Student engagement ratings are calculated by NTU's Learning Analytics system, StREAM, which reports daily on individual-level engagement data. NTU's engagement ratings are a weighted sum of seven data points: VLE logins, VLE learning room access, attendance, online submissions, online resource use, building access, and library loans. Each of these data points is a daily binary measure for which, if a student completes an action related to one of the data points, they receive a 1, and they otherwise receive a 0. The weighted sum is then combined over several days and subjected to a smoothing function, to compute a discrete daily rating of 1 to 5. Therefore, both primary outcomes reflect the average daily engagement of a student across the measurement period (short and medium term respectively).

There are two secondary outcomes:

I. Whether an additional no-engagement alert was generated in Term 1
II. Whether the student answers the support phone call (including in the model all students in intervention 1 and those students in intervention 2 who booked a phone call)[2]

There are two exploratory outcomes:

I. Whether the student answers the support phone call (including in the model all alert-generating students)
II. Attendance (the proportion of timetabled sessions attended) in the two-week period following the alert[3]

[2] This outcome was listed in the trial protocol as 'phone call takes place' but is clarified here to mean that the student answers the call.
[3] Not all sessions are monitored for attendance. Every session where the room and lecturer are booked through the central timetabling system can be subject to attendance monitoring. However, it is decided by individual departments what type of session is timetabled centrally (seminars, lectures, tutorials) and then by individual lecturers if they want to use the system, take a register and use the attendance monitoring

3.4. Sample selection

The participant pool was composed of all eligible NTU undergraduate students who received at least one no-engagement alert in StREAM in the first term of the 2022-23 academic year. StREAM identifies students at risk of non-engagement based on students' engagement with a range of institutional systems on a daily basis. The system generates a no-engagement alert when a student did not interact with any of the items included in the LA system for 10 consecutive days during term time if they were a first year student, and 14 consecutive days for students in subsequent years. A student entered our sample the first time they generated an alert.

3.5. Analytical strategy

3.5.1. Primary outcome: Short-term engagement rating

We used the following model to estimate the effects of the intervention on the primary outcome, using ordinary least squares (OLS) regression. The 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 rating (1-5)
- 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, postcode-level marker of disadvantage (IMD quintiles), academic year group, week when the intervention was delivered, entry qualification, department[4]] and
- ϵ_i is the heteroskedasticity robust residual error term.

system. We therefore expected missing data for this variable, and we observed that 10% of students with a no-engagement alert had missing attendance data.

[4] Two changes have been made to the covariates specified in trial protocol. The analysis specified in the trial protocol did not include a "department" variable, but this omission was an accident so it has been added here. The protocol included a variable for whether anyone in the individual's family had been to university, but this variable was not collected so cannot be used in the analysis. These changes have been made to the models for all outcomes.

The β_i represent the regression coefficients. β_0 gives the value of the regression intercept. β_1 gives the value of the treatment effect. β_2 is the vector of the regression coefficients for the covariates.

3.5.2. Primary outcome: Medium-term engagement rating

We used the following model to estimate the effects of the intervention on the second primary outcome, using ordinary least squares (OLS) regression. The 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 medium term engagement rating (1-5)
- 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, postcode-level marker of disadvantage (IMD quintiles), academic year group, week when the intervention was delivered, entry qualification, department] and
- ϵ_i is the heteroskedasticity robust residual error term.

The β_i represent the regression coefficients. β_0 gives the value of the regression intercept. β_1 gives the value of the treatment effect. β_2 is the vector of the regression coefficients for the covariates.

3.5.3. Secondary outcome: Additional no-engagement flag generated in Term 1

We used the following model to estimate the effects of the intervention on whether an additional no-engagement alert was generated in Term 1, using 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 an additional no-engagement alert is generated in Term 1 (1 if they have, 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) and
- X_i is a vector of pre-intervention covariates [gender, ethnicity, postcode-level marker of disadvantage (IMD quintiles), academic year group, week when the intervention was delivered, entry qualification, department].

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

3.5.4. Secondary outcome: Student answers phone call (including in the model all students in intervention 1 and those students in intervention 2 who booked a phone call)

We used the following model to estimate the effects of the intervention on the secondary outcome of whether the student answered the phone call (including in the model all students in intervention 1 and those students in intervention 2 who booked a phone call), using 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 the student answers the call (1 if they do, 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) and
- X_i is a vector of pre-intervention covariates [gender, ethnicity, postcode-level marker of disadvantage (IMD quintiles), academic year group, week when the intervention was delivered, entry qualification, department].

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

3.5.5. Exploratory outcome: Student answers phone call (including in the model all alert-generating students)

This was analysed in the same way as the secondary outcome related to students answering the phone call, except we included all students in intervention 2 who generated an alert (not just students in intervention 2 who booked a phone call).

3.5.6. Exploratory outcome: Attendance

We used the following model to estimate the effects of the intervention on the proportion of sessions attended by a student, using ordinary least squares (OLS) regression. The 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 proportion of sessions attended by the student (ranging from 0-1)
- 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, postcode-level marker of disadvantage (IMD quintiles), academic year group, week when the intervention was delivered, entry qualification, department]; and
- ϵ_i is the heteroskedasticity robust residual error term.

The β_i represent the regression coefficients. β_0 gives the value of the regression intercept. β_1 gives the value of the treatment effect. β_2 is the vector of the regression coefficients for the covariates.

4. Results

4.1. Participant flow

Table 4 presents the proportion of the randomised sample that generated a non-engagement alert and entered the analysed sample. The proportion of participants in the randomised sample that entered the analysed sample is generally balanced across both intervention groups: 7.8% in intervention 1 and 8.7% 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 analysed sample varies substantially in terms of size as it is a subset of the randomised sample that generated an alert. Attrition - due to missing outcome data - between the point of generating an alert and the point of outcome analysis was very low (1% or less, depending on the outcome).

Table 4: Summary of proportion of students that generated an at-risk flag in each intervention group (Intervention 1 = automatic phone call; Intervention 2 = email only).

		Intervention 1	Intervention 2	Total
Number of students	Randomised	13,334	13,333	26,667
	Analysed	1,045	1,162	2,207
	Proportion of students that generated a no-engagement alert	7.8%	8.7%	8.2%

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 A2) on the demographic characteristics of the randomised sample and analysed sample revealed that there are significant differences between the samples for each recorded characteristic. The analysed sample contains a larger proportion of male students, students with entry qualifications other than A-levels, ethnic minority students (particularly those who identify as Black), students from IMD quintiles 1 and 2, and some academic departments. Students in their first year of study are also overrepresented in the analytic sample, but this is expected because their alert period is shorter than students in other years (10 days vs 14 days). Nonetheless, in each case the effect size of the difference, assessed using Cramér's V, is weak.[5]

The intervention 1 and intervention 2 groups in the analysed sample are otherwise very similar, with no substantial difference in the distribution of any of the observed characteristics.

Table 5: Distribution of covariates by intervention group (Intervention 1 = automatic phone call; Intervention 2 = email only). Note, numbers in this table are reported after HESA Standard Rounding has been applied.

	Randomised sample		Analysed sample	
	Intervention 1 (N = 13,334)	Intervention 2 (N = 13,333)	Intervention 1 (N = 1,045)	Intervention 2 (N = 1,162)
Gender				
Female	7,285 (54.6%)	7,335 (55.0%)	510 (48.7%)	565 (48.6%)
Male	5,960 (44.7%)	5,915 (44.3%)	530 (50.6%)	580 (50.0%)
Other	85 (0.6%)	80 (0.6%)	5 (0.7%)	15 (1.4%)
Missing	0 (0.0%)	5 (0.0%)	0 (0.0%)	0 (0.0%)
Ethnicity				
Asian	1,475 (11.1%)	1,515 (11.4%)	135 (12.8%)	135 (11.4%)
Black	1,605 (12.0%)	1,610 (12.1%)	175 (16.9%)	195 (16.9%)

[5] 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.

Other	1,740 (13.0%)	1,695 (12.7%)	135 (12.8%)	135 (11.8%)
White	8,515 (63.9%)	8,515 (63.9%)	600 (57.4%)	695 (59.9%)
Highest entry qualification				
A-levels	7,150 (53.6%)	7,150 (53.6%)	515 (49.1%)	555 (47.7%)
Other qualifications	6,185 (46.4%)	6,185 (46.4%)	530 (50.9%)	610 (52.3%)
Year of study				
0*	125 (0.9%)	130 (1.0%)	0 (0.0%)	0 (0.0%)
1	4,790 (35.9%)	4,790 (35.9%)	485 (46.4%)	530 (45.7%)
2	4,260 (32.0%)	4,245 (31.8%)	330 (31.8%)	360 (31.2%)
3	3,655 (27.4%)	3,650 (27.4%)	215 (20.7%)	260 (22.2%)
4	500 (3.8%)	505 (3.8%)	10 (1.1%)	10 (0.9%)
5	5 (0.0%)	10 (0.1%)	0 (0.0%)	0 (0.0%)
IMD quintile				
1	1,895 (14.2%)	1,965 (14.7%)	180 (17.0%)	235 (20.1%)
2	1,850 (13.9%)	1,925 (14.4%)	185 (17.8%)	195 (16.8%)
3	2,195 (16.5%)	2,145 (16.1%)	185 (17.8%)	190 (16.4%)
4	2,495 (18.7%)	2,445 (18.3%)	175 (16.8%)	205 (17.8%)
5	3,400 (25.5%)	3,295 (24.7%)	240 (22.8%)	245 (21.1%)
Missing	1,495 (11.2%)	1,560 (11.7%)	80 (7.8%)	90 (7.7%)
Department				
School 1	1,680 (12.6%)	1,745 (13.1%)	90 (8.4%)	140 (12.0%)
School 2	755 (5.7%)	740 (5.6%)	90 (8.6%)	80 (6.7%)
School 3	1,360 (10.2%)	1,390 (10.4%)	85 (8.3%)	100 (8.6%)
School 4	480 (3.6%)	485 (3.6%)	35 (3.3%)	30 (2.6%)

School 5*	615 (4.6%)	550 (4.1%)	0 (0.0%)	0 (0.0%)
School 6	1,880 (14.1%)	1,800 (13.5%)	180 (17.3%)	200 (17.1%)
School 7	1,040 (7.8%)	1,015 (7.6%)	95 (9.2%)	90 (7.8%)
School 8	3,065 (23.0%)	3,150 (23.6%)	295 (28.1%)	325 (27.8%)
School 9	2,450 (18.4%)	2,455 (18.4%)	175 (16.7%)	200 (17.4%)

Notes: Totals do not add up to 100% due to rounding.
*Students in academic year group 0, or those from School 5, do not generate no-engagement alerts. These students were inadvertently included as part of the randomisation and are reported in this table for completeness.

Balance checks

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

We conducted balance checks for the following covariates:[7]

- Gender
- Year of study (the student's current year of 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

[6] 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).
[7] The trial protocol specified that we would also conduct balance checks on whether the student was young or mature, student opt-outs and personal tutor opt-outs. These variables were not available so these balance checks were not performed.

report. According to this benchmark, the analysed sample is not balanced on department but is well balanced on all other 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)*	0.506	0.500	0.500	0.500	0.012
Year of student (First year)*	0.464	0.499	0.457	0.498	0.014
Department** (Art & Design)	0.084	0.278	0.120	0.325	-0.117
Entry Qualification (A-levels)	0.491	0.500	0.477	0.500	0.028

Notes: N = 2,207. All variables are binary indicators, so mean averages represent proportions of the group. The parentheses indicate the category of the covariate which was used to create the binary variable (where 1 indicates that the individual is a member of that category) in the normalised difference balance check.
* Covariates used in stratified randomisation procedure.
** Balance checks for each category of department are included in Appendix A in order to improve readability of this table. They show that the sample is balanced on all the other departments.

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 intervention groups, engagement improved over time, as shown by higher medium-term versus short-term engagement scores.

The proportion of students generating an additional no-engagement alert in Term 1 was the same across both groups (41%). The mean attendance was similar across both interventions, with participants having a mean of 0.13 in intervention 1 and 0.14 in intervention 2 (i.e. students attended 13% and 14% of their timetabled sessions respectively). Students who generated an alert therefore missed the majority of their timetabled sessions in the 10 day period following the alert.

Table 7: Average outcome scores by treatment group

Outcome	Intervention 1	Intervention 2
	Mean (SD)	**Mean (SD)**
Primary: Student's short-term engagement rating	1.77 (0.74)	1.78 (0.78)
N observations	1,032	1,159
Primary: Student's medium-term engagement rating	2.24 (0.82)	2.28 (0.84)
N observations	947	1,072
	Proportion (SD)	**Proportion (SD)**
Secondary: Additional no-engagement alert generated in Term 1	0.41 (0.49)	0.41 (0.49)
N observations	1,045	1,162
Secondary: Student answers phone call (including in the model all students in intervention 1 and those students in intervention 2 who booked a phone call)	0.41 (0.49)	0.62 (0.49)
N observations	1,045	50
Exploratory: Student answers phone call (including in the model all alert-generating students)	0.41 (0.49)	0.03 (0.17)
N observations	1,045	1,162
Exploratory: Attendance	0.13 (0.21)	0.14 (0.22)
N observations	886	965

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.

4.3. Outcome of analysis

Pre-specified analysis

Table 8 presents the estimated average effects of participating in intervention 1 versus intervention 2 on the outcomes of interest (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. Figures 2-7 visualise the effects, with 95% confidence intervals displayed.

Four out of six of the estimated effects are directionally negative, and two are positive. For four of the outcomes the effect size is negligible. Note that 'Additional no-engagement alert generated in Term 1' has a negative coefficient, as this outcome measures further non-engagement, which corresponds with a beneficial effect (increasing engagement). Neither of the estimated effects on the primary outcomes 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.[8] The effects on the secondary outcome of additional no-engagement alert generated in Term 1, and exploratory outcome of attendance are both non-significant at the 5% level.

The two outcomes capturing whether the student answers the support phone call yield different results. The effect for this outcome when including in the model all students in intervention 1 and those students in intervention 2 who booked a phone call is negative and significant at the 5% level (and at the 1% level before correcting for multiple comparisons), providing evidence that intervention 2 performs better than intervention 1 (secondary analysis) in terms of the *proportion* of students who answer a phone call. However, the effect of this outcome when including in the model all at-risk students is positive and significant at the 0.1% level (exploratory analysis) demonstrating that more phone calls are answered when calls are made by default.

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	Unadjusted p-value
	Linear regression results				
Primary: Student's short-term engagement rating (n_1 = 954, n_2 = 1,069, N = 2,023)	1.776	0.022	0.031	0.029	0.485
Primary: Student's medium-term engagement rating (n_1 = 885, n_2 = 995, N = 1,880)	2.280	-0.044	0.037	-0.053	0.232
Exploratory: Attendance (n_1 = 810, n_2 = 897, N = 1,707)	0.138	-0.001	0.010	-0.006	0.892
	Logistic regression results				

[8] Note also that School 6 had an assessment period for the 1-2 weeks of Term 2 where no alerts were sent, meaning that the estimated treatment effect on the medium-term engagement rating is slightly diluted.

Secondary: Additional no-engagement alert generated in Term 1 (n_1 = 964, n_2 = 1,072, N = 2,036)	0.419	-0.089	0.107	-0.044	0.405
Secondary: Student answers phone call (including in the model all students in intervention 1 and those students in intervention 2 who booked a phone call) (n_1 = 964, n_2 = 46, N = 1,010)	0.630	-0.845*	0.326	-0.419	0.010
Exploratory: Student answers phone call (including in the model all alert-generating students) (n_1 = 964, n_2 = 1,072, N = 2,036)	0.030	3.261***	0.195	1.114	<0.001

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 engagement rating and attendance outcomes if a student withdraws before the end of the data collection period, and are missing otherwise because they have missing data on IMD quintile. p-values for the primary and secondary outcomes have been corrected using the Benjamini-Hochberg correction.

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.

The full regression tables for each of the outcomes are in Appendix B.

+ p<0.1, * p<0.05, ** p<0.01, *** p<0.001 (significance stars reflect p-values that have been adjusted for multiple comparisons)

Figures 2 through 7 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.

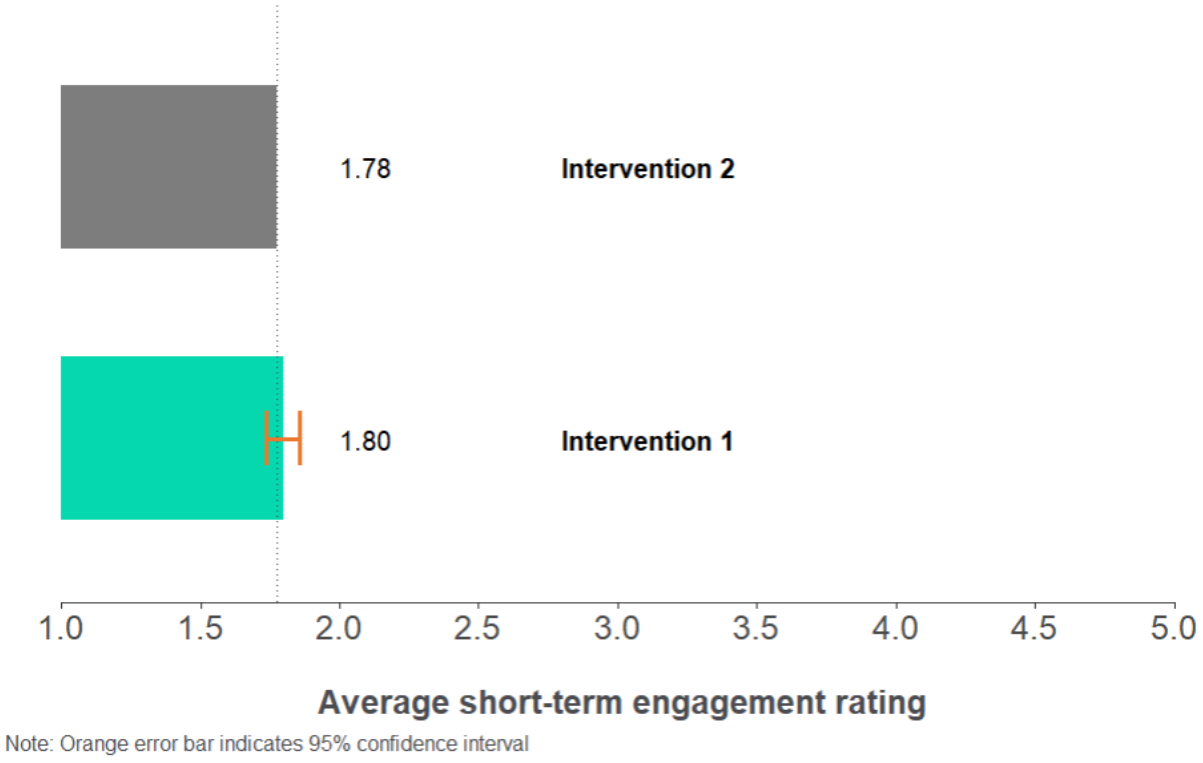


Figure 2: Average short-term engagement rating (Intervention 1 = automatic phone call; Intervention 2 = email only).

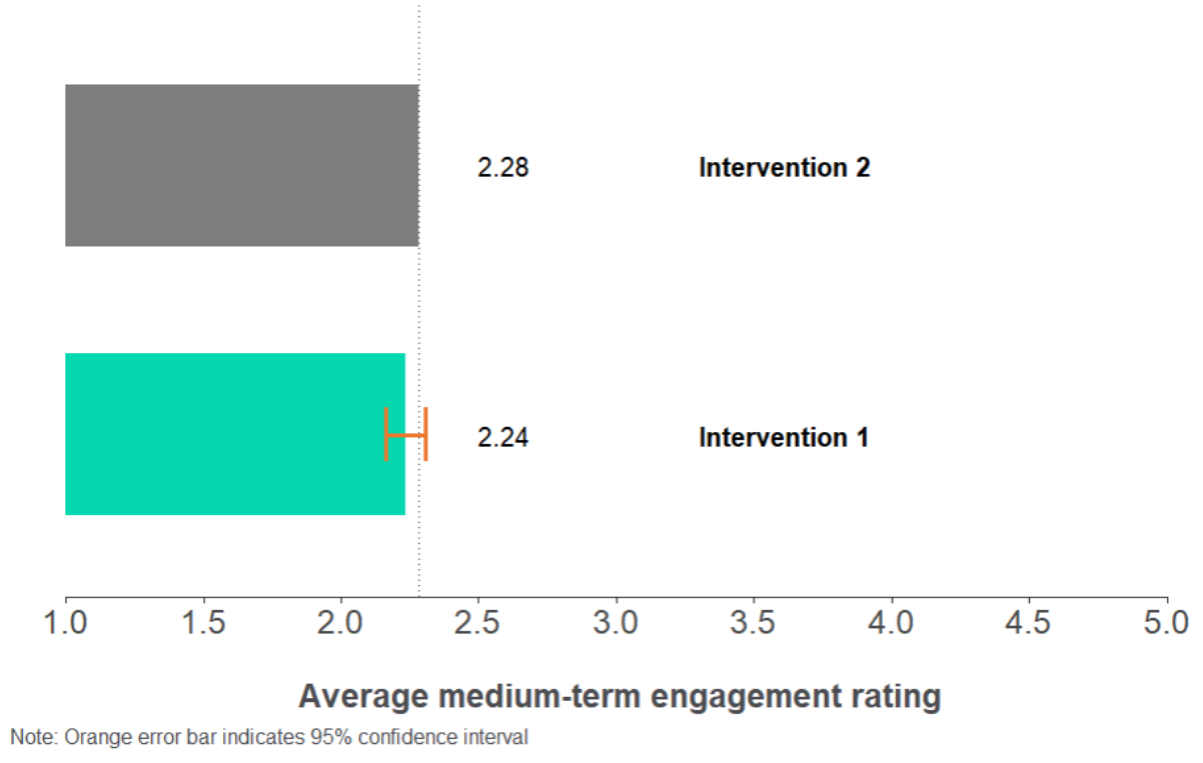


Figure 3: Average medium-term engagement rating (Intervention 1 = automatic phone call; Intervention 2

= email only).

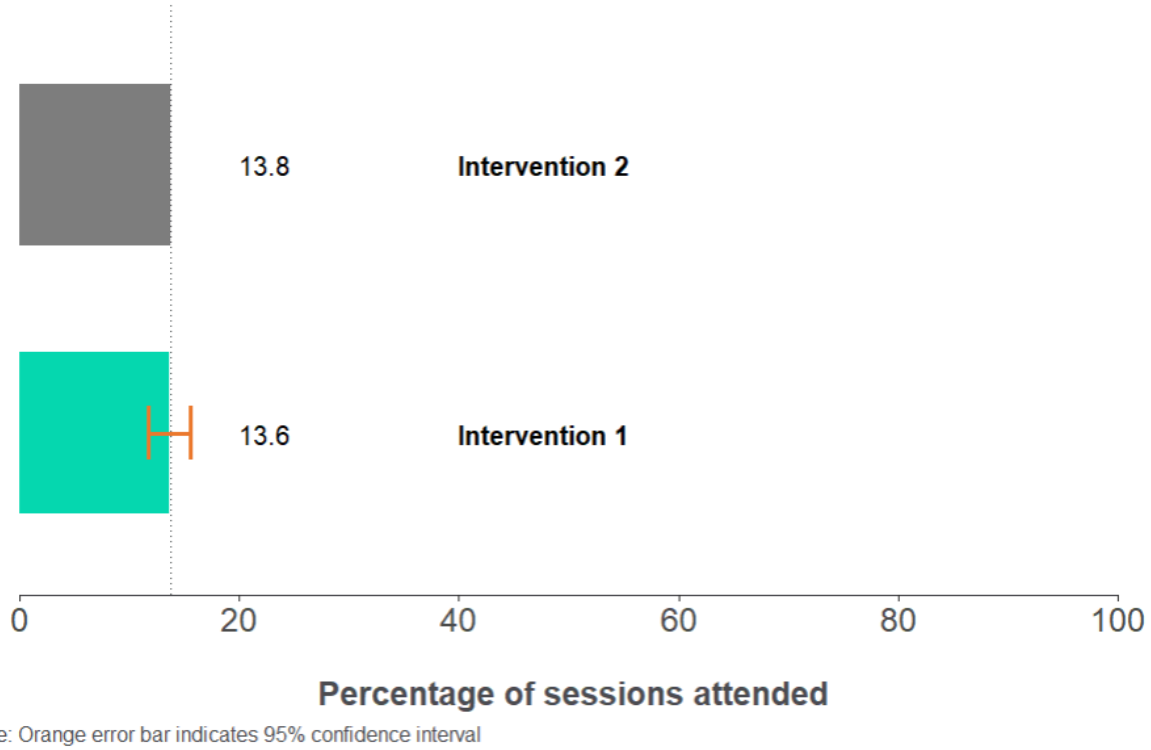


Figure 4: Average attendance in the two weeks following the intervention (Intervention 1 = automatic phone call; Intervention 2 = email only).

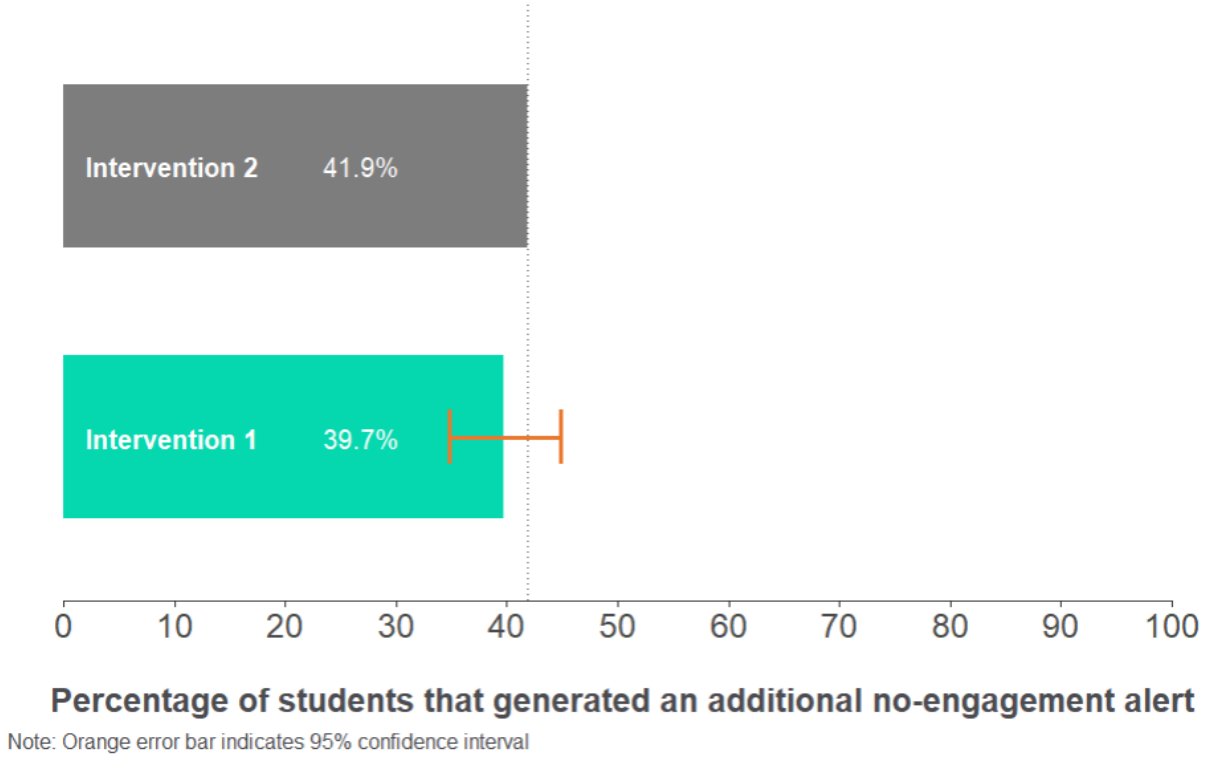


Figure 5: Percentage of students that generated an additional no-engagement alert (Intervention 1 = automatic phone call; Intervention 2 = email only).

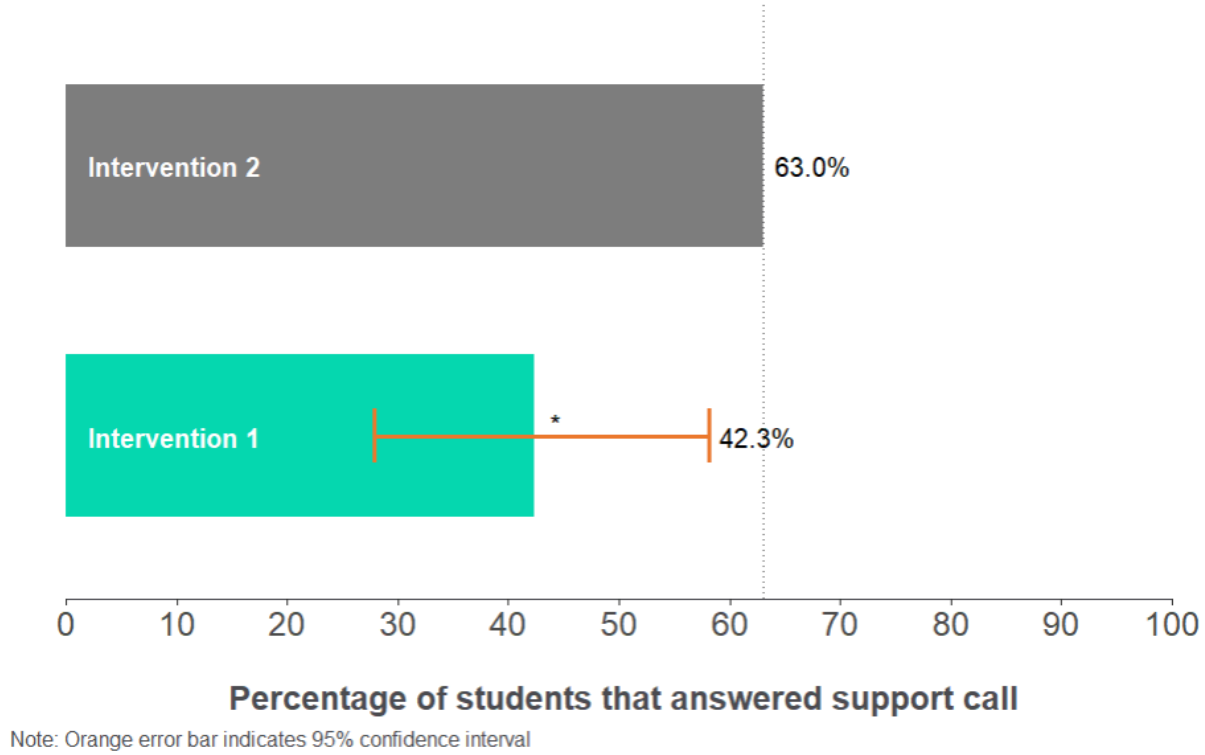


Figure 6: Percentage of students who answered the support call (including in the model all students in

intervention 1 and those students in intervention 2 who booked a phone call). * p<0.05.

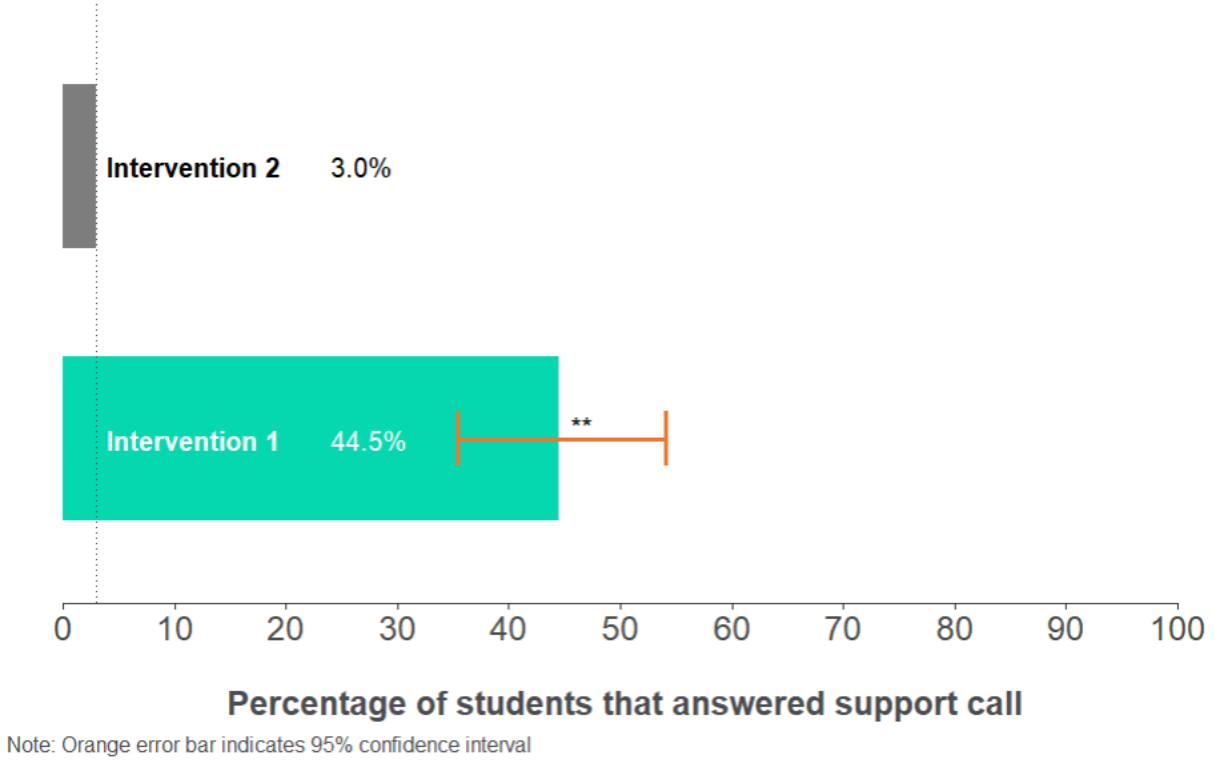


Figure 7: Percentage of students who answered the support call (including in the model all alert-generating students). ** p<0.01

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. Table 9 presents the estimated effects from the pre-specified models for each outcome, alongside the effects from the alternative models. Model 1 includes the pre-specified covariates for complete cases only; the new models (2-5) are described below.

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 to obtain the unadjusted estimates (Model 3). Both of these models produce results that closely match those of the primary analysis. Model 2 replicates the direction of the effect of intervention 1 in comparison to intervention 2 with respect to all outcomes. Model 3 replicates the direction of the effect of intervention 1 in comparison to intervention 2 with respect to all outcomes but one.

Non-compliance analysis

There was an administrative error during the trial period (on 7 November 2022) which led to members of both intervention groups being given the wrong intervention (two-sided non-compliance). To account for this non-compliance in the analysis, we have included an additional robustness check where we analyse only the compliant sample (Model 4). Model 4 (the compliant sample) replicates the direction of the effect of intervention 1 in comparison to intervention 2 with respect to all outcomes except attendance.

Matching Approach for Medium-term engagement rating

As discussed above, adjustments were made to the outcome definitions. Students in the intervention 2 group who generated two no-engagement alerts received the intervention 1 material on top of the intervention 2 material (i.e. NTU started to automatically call them). Therefore, the pre-specified average treatment effect compares students in the intervention 1 group, with a mixed group of students - some of whom received only intervention 2 as described in the trial protocol, and some who received the planned intervention 2 plus intervention 1. We refer to this mixed group as 'Intervention 2' in this report, as it captures what actually happened to students in that arm.

For the second primary outcome (students' medium-term engagement rating), the period between the generation of a first no-engagement alert and the point of outcome measurement is large enough that there are students in the intervention 2 group who received 2 or more alerts, and therefore received additional phone calls before the outcome was measured. To address this issue, we conducted two additional analyses:

1. **The pre-specified analysis, including all complete cases** (Model 1), with the caveat that the estimated treatment effect for this medium-term outcome represents a comparison of students who have received different intervention conditions than those of the short-term outcome.

2. **A matching approach** where we dropped individuals who received 2 or more no-engagement alerts from each intervention group (Model 5). This estimate is biassed in expectation as those who remain in the analytical sample do not do so randomly. This analysis may also be underpowered.

Model 5 (matching approach) replicates the direction of the effect on the student's medium-term engagement rating.

Table 9: Estimated effects (and standard errors) for the outcomes of interest 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**	**Model 5**
		Linear regression results				
Primary: Student's short-term engagement rating	Mean for intervention 2	1.776	1.782	1.782	1.743	—
	Estimated effect (SE)	0.022 (0.031)	0.017 (0.030)	-0.011 (0.032)	0.040 (0.031)	—
	N observations	2,023	2,191	2,191	1,939	—
Primary: Student's medium-term engagement rating	Mean for intervention 2	2.280	2.280	2.280	2.285	2.587
	Estimated effect (SE)	-0.044 (0.037)	-0.032 (0.036)	-0.037 (0.037)	-0.033 (0.038)	-0.064 (0.045)
	N observations	1,880	2,019	2,019	1,741	1,125
Exploratory: Attendance	Mean for intervention 2	0.138	0.142	0.142	0.130	—
	Estimated effect (SE)	-0.001 (0.010)	-0.001 (0.010)	-0.009 (0.010)	0.003 (0.010)	—
	N observations	1,707	1,831	1,831	1,627	—
		Logistic regression results				
Secondary: Additional no-engagement alert generated in Term 1	Mean for intervention 2	0.419	0.413	0.413	0.406	—
	Estimated effect (SE)	-0.089 (0.107)	-0.073 (0.103)	-0.014 (0.087)	-0.178 (0.112)	—
	N observations	2,036	2,207	2,207	1,884	—
Secondary: Student answers phone call (including in the model all students in intervention 1 and those students	Mean for intervention 2	0.630	0.620	0.620	0.667	—
	Estimated effect (SE)	-0.845* (0.326)	-0.847* (0.312)	-0.875** (0.298)	-0.993** (0.349)	—

in intervention 2 who booked a phone call)	N observations	1,010	1,095	1,095	946	—
Exploratory: Student answers phone call (including in the model all alert-generating students)	Mean for intervention 2	0.030	0.029	0.029	0.032	—
	Estimated effect (SE)	3.261*** (0.195)	3.259*** (0.189)	3.116*** (0.185)	3.208*** (0.199)	—
	N observations	2,036	2,207	2,207	1,884	—

Notes:

Model 1 = with pre-specified covariates, includes complete cases only.

Model 2 = with missing covariate data replaced with missingness indicator.

Model 3 = with no covariates.

Model 4 = compliant sample.

Model 5 = matching approach.

+ p<0.1, * p<0.05, ** p<0.01, *** p<0.001 (significance stars reflect p-values that have been adjusted for multiple comparisons)

5. Discussion

Interpretation

All pre-specified analyses show no difference between students receiving intervention 1 (automatic phone call) over intervention 2 (email), in terms of their engagement at university as measured by the learning analytics system. Estimated effects are substantively small and none are statistically significant at the 5% level, with the exception of the outcomes relating to whether the student answered the support phone call, which is significant at the 5% and 0.1% levels, depending on the denominator used.

For students who have generated no-engagement alerts, automatically calling them to offer support is no more effective in changing their engagement rating than emailing them with an offer of a call. Many more call attempts do take place if they are conducted by default (1,045) rather than booked by students in the intervention 2 group (50), but they are not shown to increase short- or medium-term engagement ratings. There is also no evidence for a difference between groups in the percentage of students who received an additional no-engagement alert (in total, 399 students in the analysis sample for intervention 1 and 449 students in intervention 2 received an additional alert).

The analysis of whether the student answers the support phone call yields different results depending on the reference group. As noted above, many more call attempts

were made to students in the intervention 1 group (automatic calling) than students in the intervention 2 group (who could choose to book a support call), but students who booked a call were significantly more likely to answer it.

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. The analysed sample differs from the wider population of students at NTU on the characteristics observed, but has, for example, a higher proportion of male students and a higher proportion of students from ethnic minority backgrounds. These differences are a reflection of the type of students who are at risk of generating a no-engagement alert. It is plausible that the effects found at NTU may generalise to other universities but the extent to which they do will likely depend in part on the similarity of their student population.

On the third type of generalisation, we know that engagement varies throughout the academic year. Drivers of disengagement and re-engagement may therefore also differ over time, as highlighted by the difference between the short-term and medium term engagement ratings. The demands placed on students vary across the year, from reading weeks to exam periods. Effects may therefore be different at different times of the year, and some interventions may be more efficient earlier or later in the academic year.

Trial limitations

For the second primary outcome (a student's medium-term engagement rating), the period between a first no-engagement alert and the point of outcome measurement is long enough that there are students in the intervention 2 group who received 2 or more alerts (and therefore some default phone calls) before the outcome was measured. The additional analyses conducted to address the issue (see Robustness Checks, above) produced almost exactly the same result as the pre-specified analysis, so we can be reasonably confident that intervention 1 is no more effective than intervention 2 in terms of a students' medium term engagement.

There was an administrative error during the trial period (on 7 November 2022) which led to members of both intervention groups being given the wrong intervention (two-sided non-compliance). To explore the effect of this issue on the results, we first

checked the balance of non-compliance across the two intervention groups and found that it is balanced across arms. To account for the non-compliance in the analysis, we have included an additional robustness check where we analyse only the compliant sample for all outcomes. Again, the results of this check are almost exactly the same as those from the primary analysis so we have good confidence in our overall conclusions.

The raw data that comprises the discrete five-point engagement rating was not available for analysis. Students who have triggered a no-engagement alert could increase their engagement post-intervention without their engagement rating moving from the lowest category. Furthermore, students within the same engagement rating may have quite different levels of engagement that would be revealed with an analysis of the raw data.

The trial duration limited the range of outcome measures to those that were available at the time of analysis and to those measured by the learning analytics system. Other longer term outcomes that could be tested are students' attainment, continuation and progression.

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.

Foster, E. and Siddle, R., 2020. The effectiveness of learning analytics for identifying at-risk students in higher education. *Assessment & Evaluation in Higher Education*, *45*(6), pp.842-854.

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

Appendix A: Balance checks

Table A1: Full table of balance checks for department.

	Intervention 1		Intervention 2		
Department	**Mean**	**(S.D.)**	**Mean**	**(S.D.)**	**Normalised difference**
School 1	0.084	0.278	0.120	0.325	-0.117
School 2	0.086	0.281	0.067	0.250	0.071
School 3	0.083	0.276	0.086	0.281	-0.010
School 4	0.033	0.180	0.026	0.159	0.045
School 6	0.173	0.379	0.171	0.377	0.005
School 7	0.092	0.289	0.078	0.269	0.049
School 8	0.281	0.450	0.278	0.448	0.008
School 9	0.167	0.373	0.174	0.379	-0.020

Notes: N = 2,207. 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 normalised difference balance check. Students from School 5 do not generate no-engagement alerts. These students were inadvertently included as part of the randomisation but are excluded from these tests.

Table A2: 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	35.49	3	<0.001	0.035
Ethnicity	48.95	3	<0.001	0.041
Highest entry qualification	22.81	1	<0.001	0.028
Year of study*	126.35	4	<0.001	0.066
IMD quintile	73.77	5	<0.001	0.051
Department*	66.63	7	<0.001	0.049

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

*Students in academic year group 0, or those from School 5, do not generate no-engagement alerts. These students were inadvertently included as part of the randomisation but are excluded from these tests.

#Interpretation of Cramérs V is dependent on the degrees of freedom (Cohen, 1988) but in all cases reported here values <0.07 indicate no 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 and likelihood of generating an additional no-engagement alert.

	Student's short-term engagement rating	Student's medium-term engagement rating	Attendance
(Intercept)	1.769	1.950	0.260
	s.e. = 0.082	s.e. = 0.100	s.e. = 0.029
	p = <0.001	p = <0.001	p = <0.001
Allocation (Ref: Intervention 2)			
Intervention 1	0.022	−0.044	-0.001
	s.e. = 0.031	s.e. = 0.037	s.e. = 0.010
	p = 0.485	p = 0.232	p = 0.892
Gender (Ref: Female)			
Male	−0.072	−0.160	-0.036
	s.e. = 0.036	s.e. = 0.043	s.e. = 0.012
	p = 0.044	p = 0.017	p = 0.003
Ethnicity (Ref: White)			
Asian	0.069	0.005	-0.023
	s.e. = 0.055	s.e. = 0.063	s.e. = 0.016
	p = 0.205	p = 0.937	p = 0.161
Black	0.018	0.056	-0.022
	s.e. = 0.042	s.e. = 0.052	s.e. = 0.013
	p = 0.675	p = 0.284	p = 0.099
Other	0.117	−0.059	-0.003

	s.e. = 0.053	s.e. = 0.060	s.e. = 0.018
	p = 0.028	p = 0.324	p = 0.885
IMD Quintile (Ref: Quintile 1)			
Quintile 2	−0.072	0.048	-0.014
	s.e. = 0.047	s.e. = 0.059	s.e. = 0.016
	p = 0.131	p = 0.416	p = 0.380
Quintile 3	−0.101	0.060	-0.014
	s.e. = 0.049	s.e. = 0.062	s.e. = 0.017
	p = 0.038	p = 0.335	p = 0.400
Quintile 4	0.059	0.101	-0.006
	s.e. = 0.051	s.e. = 0.060	s.e. = 0.016
	p = 0.248	p = 0.091	p = 0.730
Quintile 5	0.039	0.217	0.001
	s.e. = 0.050	s.e. = 0.060	s.e. = 0.017
	p = 0.437	p = <0.001	p = 0.970
Academic year (Ref: Year 1)			
Year 2	-0.272	0.045	-0.052
	s.e. = 0.035	s.e. = 0.043	s.e. = 0.011
	p = <0.001	p = 0.299	p = <0.001
Year 3	-0.153	-0.092	-0.006
	s.e. = 0.042	s.e. = 0.050	s.e. = 0.015
	p = <0.001	p = 0.067	p = 0.708
Year 4	-0.302	-0.280	0.132
	s.e. = 0.135	s.e. = 0.275	s.e. = 0.087

	p = 0.025	p = 0.309	p = 0.129
Week when intervention was delivered (Ref: Week 40 - w/c 3 Oct 2022)			
Week 41	-0.033	0.006	-0.032
	s.e. = 0.062	s.e. = 0.083	s.e. = 0.025
	p = 0.601	p = 0.944	p = 0.204
Week 42	0.005	0.015	-0.032
	s.e. = 0.061	s.e. = 0.086	s.e. = 0.026
	p = 0.939	p = 0.865	p = 0.220
Week 43	0.134	0.057	-0.025
	s.e. = 0.066	s.e. = 0.087	s.e. = 0.025
	p = 0.044	p = 0.513	p = 0.321
Week 44	0.506	0.324	0.037
	s.e. = 0.066	s.e. = 0.083	s.e. = 0.025
	p = <0.001	p = <0.001	p = 0.143
Week 45	0.482	0.363	0.050
	s.e. = 0.077	s.e. = 0.092	s.e. = 0.026
	p = <0.001	p = <0.001	p = 0.054
Week 46	0.550	0.395	0.062
	s.e. = 0.081	s.e. = 0.098	s.e. = 0.027
	p = <0.001	p = <0.001	p = 0.024
Week 47	0.254	0.300	0.011
	s.e. = 0.080	s.e. = 0.104	s.e. = 0.030

	p = 0.002	p = 0.004	p = 0.725
Week 48	0.249	0.436	-0.045
	s.e. = 0.068	s.e. = 0.088	s.e. = 0.023
	p = <0.001	p = <0.001	p = 0.052
Week 49	0.277	0.517	-0.061
	s.e. = 0.063	s.e. = 0.079	s.e. = 0.024
	p = <0.001	p = <0.001	p = 0.011
Week 50	0.093	0.221	-0.022
	s.e. = 0.170	s.e. = 0.164	s.e. = 0.068
	p = 0.584	p = 0.178	p = 0.753
Highest entry qualification (Ref: A-levels)			
Other qualifications	-0.073	-0.089	0.011
	s.e. = 0.031	s.e. = 0.038	s.e. = 0.010
	p = 0.021	p = 0.020	p = 0.271
Department (School 1)			
School 4	-0.207	-0.098	-0.043
	s.e. = 0.093	s.e. = 0.106	s.e. = 0.033
	p = 0.026	p = 0.355	p = 0.189
School 3	-0.157	0.054	-0.097
	s.e. = 0.078	s.e. = 0.088	s.e. = 0.027
	p = 0.045	p = 0.536	p = <0.001
School 1	-0.154	-0.039	-0.073
	s.e. = 0.073	s.e. = 0.082	s.e. = 0.026

	p = 0.036	p = 0.638	p = 0.006
School 6	0.078	0.215	-0.092
	s.e. = 0.070	s.e. = 0.072	s.e. = 0.023
	p = 0.267	p = 0.003	p = <0.001
School 7	-0.166	-0.086	-0.153
	s.e. = 0.074	s.e. = 0.080	s.e. = 0.021
	p = 0.025	p = 0.286	p = <0.001
School 9	-0.012	0.213	-0.053
	s.e. = 0.067	s.e. = 0.073	s.e. = 0.022
	p = 0.858	p = 0.004	p = 0.018
School 8	-0.047	0.216	-0.087
	s.e. = 0.061	s.e. = 0.065	s.e. = 0.021
	p = 0.436	p = <0.001	p = <0.001
Num.Obs.	2,023	1,880	1,707

Table B2: Full table of regression coefficients for the pre-specified models for attendance and whether or not a student answers a phone call (two models).

	Additional no-engagement alertgenerated in Term 1	Student answers phone call (all students in intervention 1 and those students in intervention 2 who booked a phone call)	Student answers phone call (all alert-generating students)
(Intercept)	0.902	0.465	-3.636
	s.e. = 0.280	s.e. = 0.465	s.e. = 0.377
	p = 0.001	p = 0.318	p = <0.001
Allocation (Ref: Intervention 2)			
Intervention 1	-0.089	-0.845	3.261

	s.e. = 0.107	s.e. = 0.326	s.e. = 0.195
	p = 0.405	p = 0.010	p = <0.001
Gender (Ref: Female)			
Male	0.244	0.036	0.083
	s.e. = 0.124	s.e. = 0.159	s.e. = 0.151
	p = 0.050	p = 0.822	p = 0.583
Ethnicity (Ref: White)			
Asian	0.002	0.145	0.148
	s.e. = 0.178	s.e. = 0.220	s.e. = 0.211
	p = 0.990	p = 0.510	p = 0.481
Black	0.262	0.708	0.702
	s.e. = 0.156	s.e. = 0.192	s.e. = 0.181
	p = 0.092	p = <0.001	p = <0.001
Other	0.107	-0.075	-0.079
	s.e. = 0.186	s.e. = 0.228	s.e. = 0.221
	p = 0.566	p = 0.741	p = 0.722
IMD Quintile (Ref: Quintile 1)			
Quintile 2	0.225	0.073	0.051
	s.e. = 0.170	s.e. = 0.214	s.e. = 0.205
	p = 0.186	p = 0.733	p = 0.803
Quintile 3	0.205	0.379	0.274
	s.e. = 0.174	s.e. = 0.224	s.e. = 0.211
	p = 0.240	p = 0.091	p = 0.194
Quintile 4	0.030	0.253	0.143
	s.e. = 0.176	s.e. = 0.227	s.e. = 0.215

	p = 0.864	p = 0.265	p = 0.506
Quintile 5	0.067	-0.002	-0.108
	s.e. = 0.172	s.e. = 0.223	s.e. = 0.214
	p = 0.696	p = 0.993	p = 0.614
Academic year (Ref: Year 1)			
Year 2	-0.764	0.261	0.244
	s.e. = 0.127	s.e. = 0.154	s.e. = 0.147
	p = <0.001	p = 0.091	p = 0.097
Year 3	-0.563	0.332	0.263
	s.e. = 0.145	s.e. = 0.181	s.e. = 0.171
	p = <0.001	p = 0.067	p = 0.125
Year 4	-1.720	-0.799	-0.846
	s.e. = 0.540	s.e. = 0.809	s.e. = 0.796
	p = 0.001	p = 0.323	p = 0.288
Week when intervention was delivered (Ref: Week 40 - w/c 3 Oct 2022)			
Week 41	-0.313	-0.068	-0.111
	s.e. = 0.200	s.e. = 0.266	s.e. = 0.256
	p = 0.117	p = 0.799	p = 0.665
Week 42	0.046	0.364	0.372
	s.e. = 0.220	s.e. = 0.295	s.e. = 0.280
	p = 0.835	p = 0.217	p = 0.184
Week 43	-0.452	-0.076	-0.044
	s.e. = 0.222	s.e. = 0.305	s.e. = 0.293
	p = 0.042	p = 0.803	p = 0.879

Week 44	-0.995	0.429	0.448
	s.e. = 0.202	s.e. = 0.276	s.e. = 0.263
	p = <0.001	p = 0.121	p = 0.089
Week 45	-1.463	-0.026	0.016
	s.e. = 0.222	s.e. = 0.298	s.e. = 0.287
	p = <0.001	p = 0.932	p = 0.956
Week 46	-1.990	-0.026	0.006
	s.e. = 0.246	s.e. = 0.325	s.e. = 0.312
	p = <0.001	p = 0.936	p = 0.983
Week 47	-1.897	0.708	0.551
	s.e. = 0.273	s.e. = 0.355	s.e. = 0.328
	p = <0.001	p = 0.046	p = 0.094
Week 48	-3.505	0.146	0.150
	s.e. = 0.325	s.e. = 0.299	s.e. = 0.286
	p = <0.001	p = 0.626	p = 0.601
Week 49	-6.371	-0.054	-0.009
	s.e. = 1.016	s.e. = 0.277	s.e. = 0.268
	p = <0.001	p = 0.847	p = 0.972
Week 50	-16.601	0.999	0.973
	s.e. = 293.01	s.e. = 0.670	s.e. = 0.602
	p = 0.955	p = 0.136	p = 0.106
Highest entry qualification (Ref: A-levels)			
Other qualifications	-0.067	-0.012	0.016
	s.e. = 0.108	s.e. = 0.136	s.e. = 0.130
	p = 0.536	p = 0.930	p = 0.903

Department (Ref: School 1)			
School 4	0.438	-1.340	-1.326
	s.e. = 0.366	s.e. = 0.485	s.e. = 0.472
	p = 0.231	p = 0.006	p = 0.005
School 3	0.239	-0.363	-0.414
	s.e. = 0.262	s.e. = 0.336	s.e. = 0.313
	p = 0.362	p = 0.280	p = 0.186
School 2	0.636	-0.576	-0.496
	s.e. = 0.265	s.e. = 0.316	s.e. = 0.301
	p = 0.016	p = 0.068	p = 0.099
School 6	-0.063	-0.559	-0.499
	s.e. = 0.231	s.e. = 0.290	s.e. = 0.271
	p = 0.783	p = 0.054	p = 0.065
School 7	0.951	-0.681	-0.614
	s.e. = 0.257	s.e. = 0.314	s.e. = 0.297
	p = <0.001	p = 0.030	p = 0.039
School 9	-0.022	-0.294	-0.310
	s.e. = 0.225	s.e. = 0.289	s.e. = 0.270
	p = 0.924	p = 0.310	p = 0.251
School 8	-0.058	-0.716	-0.605
	s.e. = 0.202	s.e. = 0.256	s.e. = 0.238
	p = 0.775	p = 0.005	p = 0.011
Num.Obs.	2,036	1,010	2,036

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*[9]	*Is it Type 1,2 or 3 evidence - according to the OfS standard of evidence?*
PRIMARY: Student's short-term engagement rating	2,023	0.485	0.029 (Hedges g)	-	3.9	3
PRIMARY: Student's medium-term engagement rating	1,880	0.232	-0.053 (Hedges g)	-	3.9	3
SECONDARY: Additional no-engagement alert generated in Term 1	2,036	0.405	-0.044 (Cohen's h)	-	3.9	3

[9] *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.*

Outcome	Sample size	P Value	Effect	Estimated 'real world' effect	Evaluation security (1 = not at all secure 5 = very secure)	Type of evidence
SECONDARY: Student answers phone call (including in the model all students in intervention 1 and those students in intervention 2 who booked a phone call)	1,010	0.010	-0.419 (Cohen's h)	-	3.9	3