A Winning Start
Statistical Analysis Plan

Evaluator (institution): NFER
Principal investigator: Ben Styles

PROJECT TITLE	A Winning Start trial (second run)
DEVELOPER (INSTITUTION)	Education Endowment Foundation (EEF)
EVALUATOR (INSTITUTION)	National Foundation for Educational Research (NFER)
PRINCIPAL INVESTIGATOR(S)	Dr Ben Styles
SAP AUTHOR(S)	Dr Joana Andrade, Dr Ben Styles, Palak Roy and Andrew Smith
TRIAL DESIGN	Two-armed randomised controlled trial (RCT) with a crossover design; Random allocation of teachers
TRIAL TYPE	Feasability of RCT
PUPIL AGE RANGE AND KEY STAGE	Ages 12-13 (Year 8, Key Stage 3)
NUMBER OF SCHOOLS	18
NUMBER OF TEACHER-CLASS UNITS	74
NUMBER OF PUPILS	1933 (estimated)[1]
PRIMARY OUTCOME MEASURE AND SOURCE	Pupils' improved learning in science as measured by the end-of-topic tests
SECONDARY OUTCOME MEASURE AND SOURCE	n/a

SAP version history

VERSION	DATE	REASON FOR REVISION
2.0 [*latest*]	7 April 2022	The trial was originally scheduled to run during the spring term of the academic year 2019-20, and was rescheduled to run again during the spring and summer terms of 2021. Although some data collection was possible, the original trial could not be completed due to Covid-19 pandemic related UK national lockdowns. The original SAP was therefore revised for the second run of the trial taking place between autumn 2021 and spring 2022. Randomisation took place in October and November 2021.
1.0 [*original*]	25 June 2020	N/A

[1]Number of pupils are estimated based on the randomised units. Some participating schools have not yet submitted pupil data hence the estimated number

Table of contents

Introduction

In 2018 EEF undertook a review of their grant-making processes, identifying that teachers and headteachers were keen to answer research questions about the everyday choices that teachers have to make when planning their lessons and deciding on the pedalogical approaches they take in their classrooms. EEF commissioned NFER to conduct two classroom based 'Teacher Choices' trials as a result, the principal aim of these being to explore the feasibility of running robust evaluations of teachers' everyday choices in the classroom.

This 'A Winning Start' trial aims to answer research questions about the feasibility of this type of trial as a research method, as well as to undertake implementation and process evauation, and to understand the impact of the teacher choice on pupils' learning. In this trial the teacher choice is a comparison of two different approaches to starting KS3 Science lessons: quizzes and discussions (the active control).

The trial research questions relevant to this SAP are as follows, and are a subset of those included in the Winning Start (*second run*) evaluation protocol (Roy et al., 2021):

- Can the end-of-topic tests be used to measure the impact of teacher choices RCTs? (RQ3)
- Is it feasible to combine or link attainment data from different teacher-made tests across a number of topics? (RQ4)
- Does a given teacher choice improve pupil attainment outcomes as measured by teacher-made tests? (RQ6)
- What can we learn from the trials about the expected effect sizes for future teacher choice trials? (RQ5)

These research questions differ to those specified in the SAP of the *first run* of the trial[2], which also asked:

[2] https://d2tic4wvo1iusb.cloudfront.net/documents/projects/A_winning_start_SAP.pdf?v=1630925274

- Is the level of teacher guidance on the 'choices' appropriate to interpret and implement the selected approach and to establish compliance with the intended 'choices'?
- If end-of-topic tests can be used to measure the impact of teacher choices RCTs, what psychometric properties do these tests need in order to be used as a valid outcome measure for future trials?
- What can we learn from the trials about intra-cluster correlations, pre-post correlations and sample sizes for future teacher choices trials?

Although the findings of the *second run* of the trial are not expected to directly answer these additional *first run* questions, the findings of both trial interations will be presented in the final evaluation report.

Design overview

Trial design, including number of arms		Two-armed randomised controlled trial (RCT) with a crossover design
Unit of randomisation		Teacher-class unit (TCU)[3]
Stratification variables		School
Primary outcome	variable	Pupils' learning in science
	measure (instrument, scale, source)	School selected end-of-topic tests; GL Progress Test in Science (PTS)
Secondary outcome(s)	variable(s)	n/a
	measure(s) (instrument, scale, source)	n/a
Baseline for primary outcome	**variable**	n/a
	measure (instrument, scale, source)	n/a
Baseline for secondary outcome	**variable**	n/a
	measure (instrument, scale, source)	n/a

Teacher-class unit (TCU) randomisation

The randomisation design described in the second run trial protocol (Roy et al., 2021) specified the randomisation of TCUs stratified by school (see Appendix A for an example of

[3] If a teacher is teaching more than one class, these are combined in the same teacher-class unit.

the R code used for randomisation). The use of TCUs was an innovation to the trial design (i.e. that specified in Roy et al., 2021) and a response to two thirds (n=12) of the participating schools deviating from the processes and procedures set out in the trial's Memorandum of Understanding (MOU). The evaluation team became aware of several instances of:

- a teacher teaching several classes when the MOU required each teacher to teach a single class;
- more than one teacher listed as teaching the same class when the MOU required that each class was taught by a single science teacher;
- more than one science topic being selected per term when the MOU required that schools selected a single topic per term;
- different topics being assigned to different participating classes when the MOU required that all the participating classes in a school were to be taught and tested on the same topics.

For a given topic within a school, TCUs were randomly allocated to either quiz or discussion starters in the second autumn half-term. They then switched to the alternate treatment in the first spring half-term. Randomisation was undertaken in three tranches in October and November 2021. The need for three separate randomisations was due to problems with data collection which did not reflect the design or set-up of the trial, specifically schools did not follow the eligbility criteria and did not supply correct teacher or class information by the cut-off date.

Randomisation Tranche	Schools	Teachers	Classes
Tranche 1 (T1)	6	21	22
Tranche 2 (T2)	9	36	41
Tranche 3 (T3)	3	11	11
T1,T2 and T3 total	18	68	74

Sample size calculations overview

A conservative approach to sample size calculation was adopted that did not assume a crossover design due to the uncertainty around calibrating different end-of-topic tests. Instead, it assumed a standard parallel-groups cluster design with the following parameters[4]:

		Protocol	Randomisation	
		OVERALL	OVERALL	FSM
Minimum Detectable Effect Size (MDES)		0.20	0.20	0.29[5]
Pre-test/ post-test correlations	level 1 (pupil)	0.70	0.70	0.70
	level 2 (teacher)			
	level 3 (school)			
	level 2 (teacher)	0.17	0.17	0.17

[4] For context, the first run trial SAP (p.5) includes further discussion of sample size assumptions. Note that the second run does not randomise near-ability sets within the same school to different trial arms.
[5] Note that the main aim of the trial is to explore the feasibility of running robust evaluations of teachers' everyday choices in the classroom. Given that the main objective is about trial feasibility, the trial is deliberately underpowered to detect an impact.

		Protocol	Randomisation	
		OVERALL	OVERALL	FSM
Intracluster correlations (ICCs)	level 3 (school)			
Alpha		0.05	0.05	0.05
Power		0.8	0.8	0.8
One-sided or two-sided?		Two-sided	Two-sided	Two-sided
Average cluster size		30	26.12	3.68*
Number of TCUs	quiz	80	74	74
	discussion	80	74	74
	total	80	74	74
Number of pupils	quiz	1187		
	discussion	1187		
	total	2374	1933	272

**14.1% of secondary school pupils were eligible for FSM in January 2019[6]. We collected FSM data directly from schools for this trial so will not be able to use everFSM.*
***Note that we may lose some clusters for the FSM analysis due to the absence of FSM-eligible children in higher science sets.*

Year 8 was chosen over Year 7 as the school routine for Year 8 students is more established and science teachers are likely to pilot new techniques in Year 8 classes before the students move closer to Key Stage 4.

Analysis

Primary outcome analysis

This trial focuses on one outcome, pupils' learning in science, as measured by school selected end of topic tests and the GL Progress Test in Science (PTS). In accordance with the EEF's 2018 statistical analysis guidance (EEF, 2018) the primary outcome analyses in this section will assume intention-to-treat.

As this trial seeks both to estimate impact and to also understand the feasibility of this type of design more generally, two primary outcome analyses will be undertaken in R, principally to investigate the feasibility of each:

1. **Cluster crossover analysis** (based on the calibration of tests across topics within each school, and assuming that calibration is possible);
2. **Meta-analysis** of individual test effect sizes.

6 https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/812539/Schools_Pupils_and_their_Characteristics_2019_Main_Text.pdf

Analysis method 1: cluster crossover analysis of linked (calibrated) data

Assuming that it is possible to derive an overall science attainment score by linking the topic tests by the methods proposed in this section, we will run a crossover analysis. This second run of the trial will not include reliability analysis as this was conducted as part of the first run.

Calibration method - linking topic tests within schools: Each student will sit two end-of-topic tests (one after delivery of each of the lesson starters) and the GL Progress Test in Science (PTS). Total scores for each participating student are collected so as to not burden teachers by requesting item-level scores, and it will therefore not be possible to link the tests using IRT. We will carry out an equipercentile equate between the total score on each topic test and the total score on PTS within such schools. The method we use will follow Kolen and Brennan (1995) Section 3.4 (post-smoothed equipercentile equating using cubic splines). Ideally, the extent of smoothing will be decided according to the example in section 3.4.1 of Kolen and Brennan, i.e. using as much smoothing as possible subject to remaining within the confidence limits of the unsmoothed equipercentile procedure. However, this may not be possible if we are having to run this procedure across many schools. In practice, it may be that we set the level of smoothing after running, say, three equates and use this level across all schools.

Analysis of cluster randomised crossover trial data: As described in section 2a) of Coe and Weidmann (2019), the science attainment score of pupil *i*, under teacher *j* of school *k* at the end of period *p* can be modelled by:

$$Y_{ijkp} = \beta_0 + \delta_k + \alpha_j + u_p + v_{jp} + \tau Z_{jp} + \epsilon_{ijp}$$
$$\alpha_j \sim N(0, \sigma_\alpha^2)$$
$$v_{jp} \sim N(0, \sigma_v^2)$$
$$\epsilon_{ijp} \sim N(0, \sigma_e^2)$$

where: δ_k are school fixed effects, α_j TCU effects, u_p period effects, v_{jp} a TCU-period interaction, τ the average treatment effect, Z_{jp} identifies if TCU *j* was allocated to retrieval practice during period *p*, and ϵ_{ijp} is the measurement error for pupil *i*, in TCU *j* in period *p*.

This proposed model is based on the hierarchical models with cluster effects treated as random analysed in Turner et al (2007). While Turner and colleagues focus on designs which study different sets of subjects in the different periods, in the 'A Winning Start' trial the same cohort of pupils will be assessed in both periods and it is necessary to account for the repeated-measures aspect of the design. This will be done by including an extra pupil level random effect in the model describing the science attainment of pupil *i*, in TCU *j* at the end of period *p*.

We will hence investigate if a teacher's use of quiz as a lesson starter as opposed to a discussion starter had an effect on pupils' mastering of the contents of a science topic as measured by the linked science assessment scores. For this purpose, a multilevel model with three nested levels (period, pupil, and teacher) will be used to account for both cluster randomisation as well as crossover repeated measures design:

$$Y_{ijkp} = \beta_0 + \beta_{1i} + \beta_{2j} + \boldsymbol{\beta_3} school_k + \boldsymbol{\beta_4} period_p + \boldsymbol{\beta_{5j}} period_p + \boldsymbol{\beta_6} starter_{jp} + \epsilon_{ijp}$$

where Y_{ijkp} is the score of pupil *i*, in TCU *j*, at the end of period *p*. $school_k$ is a dummy variable for the school attended by pupil *i* during period *p*, $period_p$ is a dummy variable for the period at the end of which pupil *i* was tested, and $starter_{jp}$ is a TCU level dummy variable for the starter of the science classes of TCU *j* during period *p*.

The dummy covariate $school_k$ will be introduced in the model to account for the school-level stratified randomisation, and the teacher-level random intercepts term β_{2j} to account for the teacher-level cluster randomisation. The pupil level random intercepts term β_{1i} is included in the model to account for the repeated-measures aspect of this crossover trial. The crossover design of the trial is reflected by the introduction in the model of the fixed effect term $\boldsymbol{\beta_4} period_p$ and the TCU-level random slopes term $\boldsymbol{\beta_{5j}} period_p$.

Although EEF's statistical analysis guidance (EEF, 2018) specifies that prior attainment should be controlled for by including a pre-randomisation baseline term in the regression models used in the analyses, this is not a commonly adopted procedure in the context of crossover designs. Conceptually crossover designs can be seen as comparing the elements of a cohort to themselves at two different periods in time, which excludes the necessity for controlling for prior attainment. As such, a pre-randomisation baseline term will not be included in the analysis model, as this data is not collected for the second run of the trial.

As this is a feasibility trial we will not include a covariate for randomisation tranche. As previously stated, the use of tranches was required due to problems with data collection, and as such we suggest using a model which would be used in future similar trials, rather than adjusting our intended model to account for this. To add another covariate would also introduce additional complexity into what is already a well-specified model.

Subgroup analyses

As per the protocol, information on FSM eligibility (FSM) was collected for the Y8 pupils participating in the trial. If the primary outcome analysis is performed assuming the existence of linked data and a crossover design, a post-hoc subgroup analysis will be performed by running the primary outcome model on the FSM subsample. This conforms with the 2018 statistical analysis guidelines (EEF 2018).

Missing data

Other than outcomes, all the information (participating pupils, TCU details, school, randomised class starter at autumn and spring half terms) related to the variables in the crossover design model has been collected prior to randomisation. As such, we do not expect missingness in terms of any variables other than linked end-of-topic assessment scores. Using an AB/BA crossover design every subject is, in turn, in the discussion or quiz group depending on period. Bias due to missingness in the outcome variable is only introduced if the information is absent in one of the periods. To prevent the introduction of bias associated with this mechanism, we will only include in the analysis the cases where the pupil has a linked attainment score in both periods.

There is the possibility of TCU-level absence from end-of-topic testing at any given period of

the trial. However, it is extremely unlikely that these absences are related to any factors associated with starter allocation, and so, removing the pupils in these TCUs from the analysis is very unlikely to introduce bias, although doing so may result in a small reduction in power..

We will report levels of missingness at the end of both periods (as well as an overall level of missingness) in a CONSORT flow diagram, but we will not investigate patterns of missingness.

Compliance

A TCU level compliance measure will be considered in the trial.
For each class *k* participating in the trial on each period, we will record the number of lessons where the teacher has used the allocated lesson starter for no more than ten minutes ($Compliant_k$) and also the total number of lessons taught ($Total_k$). This record will be obtained from a short teacher survey at the end of each half-term. From this data we will derive the following TCU level measures:

$$Compliant_{TCU} = \sum_{k\ in\ TCU} Compliant_k$$
$$Total_{TCU} = \sum_{k\ in\ TCU} Total_k$$

From those we will derive an overall measure of teacher compliance:

$$Compliance_{TCU} = \frac{Compliant_{TCU}}{Total_{TCU}}$$

We will calculate the average level of compliance amongst teachers and will also analyse the distribution of the compliance measure using histograms and boxplots.

We will also analyse compliance by period and starter (quiz or discussion). We will test whether the average level of TCU compliance differs from autumn to spring half-terms and from quiz to discussion by means of paired sample t-tests. Although we can consider TCUs as being nested into schools, our analysis will be performed assuming that school is merely a stratifier of the randomisation and not a level of the model. This justifies our choice of paired sample t-tests over multilevel models for the investigation of compliance.

Intra-cluster correlations (ICCs)

Given the analytic techniques specified in this SAP, it would not make sense to report ICCs in the specific context of this trial. Instead, we will report teacher-level ICCs across the whole sample from the crossover model, acknowledging that they will not have the same meaning as a teacher-level ICC in a conventional two-level model.

Effect size calculation

If the crossover analysis can be performed using linked (calibrated) data we will explore the possibility of computing g_{IG}, a Hedges' *g* type effect size defined in Madeyski and Kitchenham (2018, p.1994). According to Madeyski and Kitchenham the standardized g_{IG} effect is comparable to effect sizes derived for independent group designs. The calculation of g_{IG} is implemented in the R package *reproducer*.

Analysis method 2: Linear regression models followed by meta-analysis

A multilevel method to estimate starter effect sizes was originally conceived when we anticipated a sampling design that involved multi-academy trusts, each using the same test. As such a design did not come to fruition, a multilevel analysis is unlikely to be effective, as to carry out a multilevel model on each topic test would likely result in 36 models (in the eventuality that there is no test overlap between schools). Whilst it is possible to automate the running of this many models, in schools with small numbers of TCUs (in some cases as few as two), multilevel models will not be able to estimate teacher-level variance. Therefore we propose to avoid the assumptions of a multilevel model and to instead use OLS regression with Huber-White standard errors for each topic test. The resulting effect sizes and variances will then be meta-analysed.

We will be working under the assumption that effects deriving from the topics being taught in the second autumn half-term and first spring half-term (period effects) as well as carryover effects[7] at the spring half-term can be regarded as negligible, and as such multilevel and OLS models are comparable.

The OLS regression will be given by the general formula:

$$\mathrm{Y} = \beta_0 + \beta_1 starter + \beta_2 TCU + \varepsilon$$

Where *Y* is the end of topic test score, β_0 is the intercept, $starter$ (β_1 estimated with Huber-White standard errors) is the allocated class starter, and $\beta_2 TCU$ is the teacher fixed effect.

Therefore we will estimate one effect per test, with the summary effect being calculated by using a fixed-effect meta-analysis of the individual models. The meta-analysis will follow the methodology prescribed in Chapter 23 of Borenstein et al (2009):

The combined effect will be calculated by creating a weighted mean effect size:

$$M = \sum_{i=1}^{k} W_i Z_i \Big/ \sum_{i=1}^{k} W_i$$

with variance:

$$V_M = 1 \Big/ \sum_{i=1}^{k} W_i$$

[7] Carryover effects correspond to effects that "carry over" from one experimental condition to another, in this case from the second autumn half-term to the first spring half-term.

where Z_i is the effect size from the ith topic test (denoted as Y in our OLS formula above), $W_i = \frac{1}{V_{Zi}}$ and V_{Zi} in the variance from the i[th] model. The effect size Z_i from each topic test OLS model will be based on the following formula:

$$Z_i = \frac{\beta_1}{\sqrt{(\widehat{var}(Y_i)}}$$

where $\sqrt{(\widehat{var}(Y_i)}$ is from a model without covariates (i.e. unadjusted). The variance V_{Zi} will be the robust standard error of β_1 squared then divided by $\sqrt{(\widehat{var}(Y_i)}$.

References

Borenstein, M., Hedges, L.V., Higgins J.P.T., and Rothstein, H.R. (2009) Introduction to Meta-Analysis, John Wiley & Sons.

Coe, R. and Weidmann, B. (2019) Design and Analysis Options, Education Endowment Froundation (September 2019).

EEF (2018) Statistical analysis guidance for EEF evaluations (March 2018).

Kolen, J.M. and Brennan, R.L. (1995) Test Equating, Scaling and Linking; Methods and Practices, Second Edition, Springer.

Madeyski, L. and Kitchenham, B. (2018) Effect sizes and their variance for AB/BA crossover design studies, Empirical Software Engineering (2018) 23, 1982-2017.

Roy, P., Harland, J., and Galvis, M. (2021) Teacher Choices trial: A Winning Start (second run) Proposed Evaluation Protocol, Education Endowment Foundation.

Turner, R.M., White, I.R., and Croudace, T. (2007). Analysis of cluster randomized cross-over trial data: a comparison of methods. Statistics in Medicine, 26(2), 274-289.

Appendix A: Example of R code used for randomisation

```
## School randomisation for EETD 2021-2022.

###Stratified by school (no other stratifiers needed)

###Last updated by Joana Andrade

###last updated on 14/10/2021

#1. Set work directory

setwd("K:/EETD/CfS/03.Randomisation")

#2.identify project

project="EETD_2021-2022"

#3.identify classification: c, r or p

classification="R"

#4. Number of the randomisation: 1st, 2nd, 3rd ...

randomisation=1

randomisation=as.character(as.roman(randomisation))

###5. Load data

Experiment<-read.csv("EETD_Teacher_Data_C.csv",header=T,stringsAsFactors = F)

Experiment$order=1:nrow(Experiment)

###Do some data cleaning and transformation before the randomisation-some teachers/classes are to be excluded

### Multiple teachers for one class can be identified by the word name being present in the Comment variable

###we have three instances of those

aux=grep("name",Experiment$Comment)

Experiment0=Experiment[aux,]

Experiment=Experiment[-aux,]

##Check that in all schools there is one topic being taught by  term with no swaps

aux=table(Experiment$NFERID,Experiment$Planned.science.topic.for.this.research.project.to.be.taught.in.Autumn.2..After.October.half.term.to.end.of.term.)

aux=as.data.frame(aux)

aux=aux[aux$Freq>0,]
```

```
aux1=aux$Var1[duplicated(aux$Var1)]
aux1=aux[aux$Var1%in%aux1,]
###Potential problems with schools 3027 and 3028. These are schools teaching more than one topic per term
aux1=Experiment[Experiment$NFERID %in%c(3028,3027),]
aux=Experiment$NFERID %in%c(3028,3027)
Experiment0=rbind(Experiment0,Experiment[aux,])
Experiment=Experiment[!aux,]

#### I'll randomise to topic in Autumn1 with school as stratifier
###sanity check: school names and numbers are 1 to 1
aux=as.data.frame(table(Experiment$NFERID,Experiment$school.name))
aux=aux[aux$Freq>0,]
sum(duplicated(aux$Var1))==0
sum(duplicated(aux$Var2))==0

Exp=Experiment[,c("NFERID","Science.teacher.name.")]
aux=colnames(Exp)
Exp=as.data.frame(table(Exp))
Exp=Exp[Exp$Freq>0,aux]

###Identify stratification and randomisation variables

#6.list the stratification variables
aux<-colnames(Exp)
stratification<-list(aux[1])
n_strats<-length(stratification)

#4.identify the randomisation/cluster variable
cluster<-aux[2]

###5. What time is now? (hh.mm)
time_now<-22.30

aux<-100*trunc(time_now)+100*(time_now-trunc(time_now))
set.seed(aux)
seeds<-sample(1:9999,size=(n_strats+2))
```

```
#Duplicated cluster information and lines with no cluster identification
#must removed
#No lines were removed
Exp<-Exp[!duplicated(Exp[cluster]),]
Exp<-Exp[!is.na(Exp[cluster]),]

#Keep the original order of the columns
originalColOrder<-colnames(Exp)

###Adding a variable that will allow for the recovery
##of the original order of the data frame rows later on
Exp$originalRowOrd<-1:nrow(Exp)

### Ordering Experiment by cluster
Exp<-Exp[order(Exp[cluster]),]

### Assigning a random order to the stratification
rands<-paste("rand",as.character(1:n_strats),sep="_")

for (i in 1:n_strats){

  aux<-as.data.frame(sort(unique(Exp[,stratification[[i]]])))
  set.seed(seeds[1])
  seeds<-seeds[-1]

  aux[rands[i]]<-sample(1:nrow(aux))

  Exp<-merge(Exp,aux,by.x=stratification[[i]],by.y=colnames(aux)[1])
}
###Randomise by cluster/randomisation variable
set.seed(seeds[1])
seeds<-seeds[-1]
Exp["rand_cluster"]<-sample(nrow(Exp))

###Reorder the rows of Experiment by rands and rancluster
rands<-c(rands,"rand_cluster")
aux<-do.call(order,Exp[rands])
Exp<-Exp[aux,]
```

```
###Assigning Control or Intervention Group
Exp$grp<-1+1:nrow(Exp)%%2

rands<-c(rands,"grp")

aux<-data.frame(starterAutumn=c("Quiz","Discussion"))
set.seed(seeds[1])
aux$randgroup<-sample(1:2)

Exp<-merge(Exp,aux,by.x="grp",by.y="randgroup")

##Returning the data frame to its original order
Exp<-Exp[order(Exp$originalRowOrd),]

###Removing the variables that are no longer necessary
rands<-c("originalRowOrd",rands)
rands<-which(colnames(Exp)%in%rands)
Exp<-Exp[,-rands]
originalColOrder=c(colnames(Experiment),colnames(Exp)[3],"starterSpring")
Experiment=merge(Experiment,Exp,all.x=TRUE)
table(Experiment$starterAutumn,useNA = "ifany")
Experiment$starterSpring="Discussion"
Experiment$starterSpring[Experiment$starterAutumn=="Discussion"]="Quiz"
table(Experiment$starterAutumn,Experiment$starterSpring,useNA = "ifany")
Experiment0$starterAutumn="not assigned"
Experiment0$starterSpring="not assigned"
Experiment0=Experiment0[,originalColOrder]
Experiment=rbind(Experiment,Experiment0)
Experiment=Experiment[order(Experiment$order),]
Experiment=Experiment[,-grep("order",colnames(Experiment))]
table(Experiment$starterAutumn,Experiment$starterSpring,useNA ="ifany")

###stopped here

###Put it out: type of document to be decided later
csvname<-
paste(paste(paste(paste(project,"Randomisation",sep="_"),randomisation,sep=""),classification,sep="
_"),
```

```
            "csv",sep=".")
write.csv(Experiment,csvname,row.names =F)
```