Implementing Randomised Control Trials in Open and Distance Learning: A Feasibility Study

Abstract

Randomised Control Trials (RCTs) are an evidence-based research approach which has not yet been adopted and widely used in open and distance education to inform educational policy and practice. Despite the challenges entailed in their application, RCTs hold the power to robustly evaluate the effects of educational interventions in distance learning and conclude on whether (or not) these interventions should be adopted and used extensively. The aim of this paper is to spark discussions around the use of RCTs in distance learning by illustrating their benefits and drawbacks including challenges in adopting RCTs in education. To achieve this aim, a RCT was implemented to examine whether a small-scale intervention in four language modules could improve attendance at an end-of-module speaking assessment, and in consequence, performance, completion and pass rates. Results raise the need for further research in order to identify what type of interventions should be designed and put into practice to elicit a positive impact on learners. The paper concludes with a discussion on why RCTs should be brought to the forefront as a viable method for the effective evaluation of the impact of open learning analytic interventions.

Keywords: Randomised Control Trials; distance education; evaluation; learning analytics

Implementing Randomised Control Trials in Open and Distance Learning: A Feasibility Study

Introduction

In a dawn of big data and applications of learning analytics in education in particular, distance learning institutions are provided with renewed opportunities to increase performance and retention and provide personalised learning on a large scale (Bienkowski, Feng, & Means, 2012; Hickey, Kelley, & Shen, 2014; Tempelaar, Rienties, & Giesbers, 2015; Tobarra, Robles-Gómez, Ros, Hernández, & Caminero, 2014). The availability of large datasets, powerful analytics engines (Tobarra et al., 2014), and proficiently designed visualisations of analytics results (González-Torres, García-Peñalvo, & Therón, 2013) may enable distance learning institutions to build on previous experiences and create insightful models of primary and perhaps real-time learning processes (Baker, 2010; Ferguson & Buckingham Shum, 2012; Papamitsiou & Economides, 2014). It is envisioned that in the very near future personalisation will be mainstream in education (Bienkowski et al. 2012, p. 5). The power of learning analytics could be used in this respect to record, visualize, and respond to learners' individual needs promptly. Yet, the impact of such learning analytics interventions need to be evaluated thoroughly to establish their effects on distance learning and personalisation, and make the most of their potential to support online learning processes. The aim of this paper is to showcase how evidence-based research, in particular Randomised Controlled Trials (RCTs) is a viable approach to testing and concluding on the effectiveness of such interventions in open and distance learning. Many open and distance learning providers seem to have high hopes that such approaches using quasi-experimental, A/B testing and RCTs can make their organisations fit-for-purpose, flexible, and innovative (Rienties, Cross, & Zdrahal, 2016). Evidence-based research could promote openness in education and support social equality by ensuring that proposed learning interventions are impactful and in support of diverse learning needs and engagement.

Experimental scientific methods are a basic approach of evidence-based research used in education to inform educational decision making (MacNeill, Campbell, & Hawksey, 2014; Slavin, 2002; 2008). Rigorous evaluations using randomised experiments (Torgerson & Torgerson, 2008) and A/B testing (Siroker & Koomen, 2013) have led to an unprecedented innovation in the last fifty years (Slavin, 2002) in fields such as medicine and agriculture. Yet, evidence-based research principles have not been adopted by educational research and learning analytics *en masse* (Hess & Saxberg, 2013; McMillan & Schumacher, 2014; Torgerson & Torgerson, 2008).

There might be valid reasons to design, implement and evaluate RCTs in distance education. For example, in some level 1 modules at the Open University UK (OU), pass rates are well below 50% (Rienties, Boroowa, et al., 2016; Rienties, Cross, & Zdrahal, 2016; Simpson, 2013). If in medicine a particular straightforward medical procedure would only be successful for half of its treated patients, or if in agriculture only half of a standard crop of wheat was harvestable, substantial review, evaluation and, where needed, investments in RCT interventions would be conducted in order to increase respective success rates. At the OU and other institutions substantial investments are made to address these retention gaps (Rienties, Cross, & Zdrahal, 2016; Calvert, 2014), but most of these interventions are done without the introduction of comparison or control groups (i.e., giving all the at-risk students exactly the same intervention) to compare the effects of the intervention(s) and randomisation of conditions.

At most distance learning providers in Europe, and the OU in particular, the appetite for conducting evidence-based interventions using RCTs is mostly limited, which in part is reflected in this journal given that only two published studies used principles of RCTs (Inkelaar & Simpson, 2015; Woods & Keeler, 2001). In part this is also a wider reflection of the education sector, whereby several ethical, fairness, social and financial concerns have been raised by academics, teachers, and students alike to implement RCTs (MacNeill et al., 2014; Slavin, 2002; 2008). A notable exception seems to be Kaplan University in the USA, where a consistent approach has been implemented to support RCT experiments on a large scale to optimise learning and teaching (Hess & Saxberg, 2013).

In addition to feelings of personal failure and lack of self-worth (Christie, Munro, & Fisher, 2004; Franssen & Nijhuis, 2011) in not completing an introductory course at an open distance programme, it is well known from marketing research (Richins, 1983; Wetzer, Zeelenberg, & Pieters, 2007) that every negative experience with a particular product or service is shared with several peers, or so-called negative word-of-mouth. In an increasingly competitive market of distance learning, it is essential that a positive brand reputation is maintained, which can be tarnished if consistently many “new” adult learners continuously fail at the first hurdle of distance education provision (Rienties, Boroowa, et al., 2016). In other words, the true “costs” of non-completion are substantially larger than merely the loss of revenue of a student deregistering (Simpson, 2013).

In this article, we will use a feasibility study of implementing a small-scale RCT to remind students to register for their final speaking exam in a range of language modules. We aim to illustrate the advantages and disadvantages of using RCTs, the affordances and feasibilities of RCTs, the financial impact and (any) ethical concerns. We hope to encourage a debate amongst academics, teachers, and educational providers as to why RCTs are so uncommon in distance education, and why there seems to be substantial resistance to such approaches despite overwhelming evidence in other disciplines and increasing evidence in other fields in education (Aboalshamat, Hou, & Strodl, 2015; Hattie, 2009; Hommes et al., 2014; Ilic & Maloney, 2014) that conducting evidence-based research is essential for moving the educational field forwards.

RCTs in Education

While in medical science, agriculture and technology conducting RCT studies has become the standard research approach in the last 50 years (Slavin, 2002; 2008), there may be some valid reasons explaining why RCT studies in education have until recently been sparse. It is widely documented that action-research, descriptive, and/or correlation studies have substantial merits for unpacking the processes and outcomes of learning. Yet, a major issue of concern in these studies is the potential of generalization of the findings beyond the context in which a particular study has been conducted (Arbaugh, 2014; Tempelaar, Rienties, & Giesbers, 2015; Hattie, 2009). Of special concern is the issue of selection bias. Even in the field of the state-of-the-art learning analytics most studies have analysed a single module or sole discipline by utilizing self-selected samples, namely teachers and/or learners who are willing to share their practices and take part in research (Papamitsiou & Economides, 2014). Such research designs are useful as they can provide insights about data analysis techniques and test the proof of concepts of data visualisation tools.

Evidence of impact is rather hard to harness without random assignment of learners and/or teachers into two or more conditions (McMillan & Schumacher, 2014; Slavin, 2008; Torgerson & Torgerson, 2008). Context specificities might alter the outcomes of an evaluation as, for example, an enthusiastic teacher might facilitate a follow-up response from his/her students on a suggested intervention while a different teacher or group of students might bring up completely different results. Likewise, in a non RCT pre-post-test design intervention the lack of comparisons with another tool or approach might lead to the identification of a positive effect of the use of the tool, which might not be reliable but rather attributed to other elements of the learning design.

An important advantage of introducing RCTs in distance education is the ability to empirically test the positive or negative effect(s) of the experimental condition(s) relative to the control condition (Siroker & Koomen, 2013; Slavin, 2008; Torgerson & Torgerson, 2008). Not every intervention will always have positive effects and may actually have unexpected, indirect or even negative effects. At present, two of the authors of this article work with an extensive number of 90+ module teams at the OU to use insights from learning analytics to help module teams to actively intervene in presentation, as well as to intervene in the learning design for the next implementation (as published in Rienties, Cross, & Zdrahal, 2016; Rienties & Toetenel, 2016). A personal observation from the authors is that several module teams seem quite willing to make a mix of interventions to address issues identified by learning analytics of real-time user data, but hardly any of the module teams seems willing to do this in an RCT manner. When asked why module teams would be unwilling to use RCTs, a typical answer given is that "it would be unethical to give the control condition an inferior treatment" or "it would be very complex to run and implement a comparison study".

This anecdotal evidence points to ethical issues related to the implementation of RCTs in education. Some teachers and researches indicate that it is rather unethical to give the control group an inferior treatment (Ziliak & Teather-Posadas, 2014). Yet, the often implicit, initial assumption that a control group receives an inferior treatment can be contested after running a RCT; it might be that the actual intervention can have a positive, neutral, or even negative effect on learning processes and retention. While there may be many theoretical and practical reasons why a particular experimental treatment might be beneficial, in real life the conditions for the control group might actually be fairly similar or even better. Implementing new interventions without appropriate RCTs might actually be "unethical", as one cannot just assume that an intervention for all will be beneficial.

For example, in a pre-post test RCT lab study (Rienties, Giesbers, Lygo-Baker, Ma, & Rees, 2016) of 36 academics who had to complete five authentic tasks using a new and unknown Virtual Learning Environment (VLE) in two conditions (experimental condition: advanced training materials with rich videos and illustration in addition to online help sheets; control condition: online help sheets only) indicated that academics in the experimental condition were substantially less satisfied with the training materials, in spite of the higher quality and support mechanisms. In addition, an unexpected side-effect of the experimental condition was that participants took 36% longer to complete a respective task, even though the advanced training materials were supposed to make it easier for academics to finish the respective task in the VLE (Rienties, Giesbers, Lygo-Baker, Ma, & Rees, 2016). Similarly, contrary to initial expectations in a small RCT design (n= 40) with three levels of frequency of audio feedback, Woods and Keeler (2001) found that both the control group and the experimental group with the least frequency of audio feedback contributed more to group discussions that the two conditions with frequent feedback. Likewise, a systematic review of nine RCTs (Ilic & Maloney, 2014) that examined the methods of teaching medical trainees evidence-based medicine identified neutral effects of RCTs. In particular, trainees' competencies in evidence-based medicine increased post-intervention in all the studies under review. However, no differences were observed between experimental and control conditions across a variety of teaching methods, such as lecture versus online and direct versus self-directed approaches (Ilic & Maloney, 2014).

Most RCTs have to go through substantial ethics review processes, whereby it is unlikely that ethics approval will be provided to a study where students are put at a substantial (negative) treatment. Second, if during the study it indeed becomes apparent that the experimental condition has structural advantages in terms of learning processes and outcomes, teachers may opt for switching the control group to the experimental treatment if the same positive effects are found after switching treatment. This might provide further evidence of the relative effectiveness of the treatment. In contrast, in a non-RCT setting a potentially inferior treatment might be continued as there is no comparison condition that can “illustrate” the ineffective design.

Reasons why RCTs are not widely used

Critics of RCTs (Ziliak & Teather-Posadas, 2014) suggest that while RCTs may be useful for lab studies, in real-life implementing large-scale RCTs in practice is complex, perhaps unethical, costly, and cumbersome as students will quickly become aware of which condition they are enrolled in, leading to known Hawthorne (i.e., being observed effect) or John Henry effect (i.e., participants in the control group are aware that they are in the disadvantaged condition and work harder to offset any potential negative effect). Indeed, a crucial element of RCTs is that participants are randomly assigned to a particular condition, as this eliminates selection bias and confounding effects (Torgerson & Torgerson, 2008). Ideally, randomisation is also blinded by those who implement and evaluate the intervention, thereby potentially offsetting Hawthorne and John Henry effects.

Moreover, it should be noted that, in the field of education, RCTs might not be adequate for measuring the impact of an intervention. Additional sources of data are often needed to fully understand and explain any impact on learning. RCTs in education might provide limited details as to why an intervention has a positive or negative impact or whether certain aspects of the intervention might be more or less effective than others. To overcome these challenges and manage to uncover how and why interventions work, RCTs should ideally be coupled with qualitative accounts of data collection. A mixed-methods approach (a mixture of qualitative and quantitative research approaches) could help address issues that relate to the complexity of the teaching and learning processes such as how faithfully an intervention has been applied to different learning contexts and how engaged teachers and/or students have been with it. This data could inform and enhance the reliability of using RCTs in education.

Apart from practical and ethical considerations, another reason that might explain why RCTs have not yet been extensively used in education might be stakeholders' resistance towards change. As Piderit (2000, p. 783) explains "[s]uccessful organisational adaptation is increasingly reliant on generating employee support and enthusiasm for proposed changes, rather than merely overcoming resistance". Resistance is often explained by a multiplicity of factors mostly related to the individuals to whom change is targeted. A number of studies tend "to blame the individual academic and attribute delays or failure in implementation to an oversimplification of negative attributes, ill-will, indolence, ineptitude or indiscipline on the part of those at whom the change is aimed … or to portray resistance to change as irrational" (Hanson 2009, p. 557),

Resistance or ambivalence could be understood in relation to attitudes. In particular, Piderit (2000) explains employees' - including academics' - ambivalence towards a change in terms of three dimensions of attitudes: cognitive, emotional and intentional. The *cognitive dimension* of attitudes refers to what academics believe about conducting RCTs. The *emotional dimension* refers to their emotions in response to RCTs. The *intentional dimension* refers to academics' intention to take some action. Piderit (2000) points out that positive and negative cognitions, feelings and intentions might co-exist. For instance, an academic might value the organisational and financial benefits of running and implementing a RCT of e-bootcamps before an exam (Rienties, Cross, & Zdrahal, 2016) for a sub-cohort of students identified as at-risk of not attending the final exam (e.g., experimental condition 1: 2 e-bootcamps of 1 hour using web video conference; experimental condition 2: 1 e-bootcamp of 2 hours using web video conference; control group: no e-bootcamp but with supportive reading materials of key aspects to prepare for the exam). Yet, their inability to control which students are enrolled in which condition, anxiety towards the appropriateness of the control condition, or perhaps the need for active interventions and uncertainty of outcomes might lead to negative emotions. Negative emotions may result in influencing cognitive attitudes and specifically the intention to pro-actively participate in RCT designs.

RCT studies in education

While not yet widely adopted and used in education and in particular distance learning education, RCTs are a topic of interest for educational stakeholders. Major funding bodies in the UK such as the Educational Endowment Foundation have raised considerably large amounts of money (£125m) to implement RCTs involving more than 2,000 schools. Similarly, the MIT's Poverty Action Lab in the US plans to conduct 500 RCTs in the field of education (Harford, 2014).

In terms of large-scale RCT interventions that have been successfully implemented in higher education, the longitudinal study of Hommes et al. (2014) provides some relevant and interesting findings that even small changes in enrolment procedures can have a substantial impact on learning. At Maastricht University, students are typically enrolled on every module at random in a new group of 10 students to work on Problem-Based Learning tasks. In a two-year study of 400+ medical students, each of the two experimental groups of 50 students were re-enrolled in subsequent modules within their respective subgroup of 50 students rather than the entire cohort of 400. Longitudinal social network analysis and psychometric analyses indicated that the experimental groups developed stronger and more learning links within their experimental group relative to the control group, and had higher (perceived) social cohesion, psychological safety, and better group dynamics (Hommes et al., 2014).

Another successfully implemented RCT in higher education (Aboalshamat et al., 2015) was implemented with more than 400 undergraduate medical and dental students. The study aimed to evaluate the effectiveness of a self-development coaching programme on the psychological health and academic performance of students. The control group attended an active placebo programme focussing on theoretical information. The experimental group attending the self-development coaching programme was found to show only short-term improvements in relation to depression and anxiety compared to the control group. The intervention had no effect on academic performance.

Building on the study of Woods and Keeler (2001), in a rather pragmatic study amongst 3374 distance learning students at University of London (Inkelaar & Simpson, 2015), approximately half of the students received a range of motivational emails to encourage them to continue studying, while the control group did not receive any motivational emails. The experimental group had 2% higher pass rates than the control condition, whereby Inkelaar and Simpson (2015) indicate that the costs of the intervention (i.e., £1000 for using system of motivational emails) was substantially lower than the increased revenue from 2% more final exam registrations (i.e., £7500).

A RCT framework in practice

The Analytics4Action Evaluation Framework (A4AEF) is a practical example of how to implement more evidence-based research in distance learning that directly involves both academics and others. The A4AEF is an evidence-based learning analytics framework currently developed at the OU with 90+ modules across the various disciplines. Among others, the A4AEF makes extensive use of RCTs to evaluate the impact of certain learning analytics interventions which aim to improve students' retention rates and performance. Figure 1 illustrates the A4AEF and how students, researchers, educators, and policy makers can evaluate and decide on the types of interventions that work well and the conditions under which this can be achieved. According to Rienties, Cross, & Zdrahal (2016) an effective evidence-based framework in learning analytics should adhere to the following conditions:

1) identify at-risk learners, or learners who need support accurately and reliably;
2) identify the improvements to be made to the learning design;
3) suggest interventions that work for both the student and the teacher, if possible, personalised ones;
4) function well in the existing teaching and learning culture; and
5) be cost-effective.

As proposed in the A4AEF, the evidence-based intervention process consists of six key steps. In step 1: key metrics and drill downs, involved stakeholders of a given module (e.g. teachers, learning analysts, administrators) gather together throughout the presentation of the module to examine learning analytics data from VLE and other systems. The aim is to understand the dynamics of the students' learning journey, unpack and translate raw data. In step 2: menu of response actions, drawing from the outcomes of step 1, a number of options on how to improve the learning design of the module are presented to involved stakeholders. The range of options available to enhance learning design might be broad yet only some of these will be feasible within the confines set by, for example, cost, practicality, and available staff time. In step 3: menu of protocols, teachers decide on how to examine the impact of the selected intervention strategy. For this purpose, they can choose from five protocols:

1) apply intervention to all students in the cohort;
2) quasi-experimental design;
3) pilot study with sub-sample;
4) A/B testing; and
5) RCTs.

In step 4: outcome analysis and evaluation, teachers determine the impact of the respective intervention in relation to specific key variables determined prior to the intervention. These might be certain learning activities, learning processes, and learning outcomes. In step 5: evidence hub, outcomes of the evaluation are shared at an institutional level through channels such as searchable repositories that are available to all institutional staff. In the final step: deep dive analysis and strategic insights, a comparison amongst different interventions across a range of modules from various disciplines is made. Strategic insights as to which interventions and under which conditions have positive impact on students' performance and retention become available. This information can lead to improvements to key metrics and the menu of response actions.

➔ Insert Figure 1 about here

Feasibility RCT study to encourage final exam registration

Simpson (2004); (2013) indicates a number of key points at which students drop out. These points are: before the first tutor marked assessment (TMA), as the learning curve becomes steeper and before the final exam/end of module assignment. Recent OU research on non-completers (IET Student Statistics and Survey Team, 2014) identifies three main reasons for withdrawing at this late stage: inability to get the time off to do the exam/assignment, not wanting to do it, and not feeling sufficiently prepared. The focus of this study is on students studying level 1 intermediate language modules in French, German, Spanish and Italian at the OU, UK. The requirements for completing these language modules is the submission of four assignments throughout the module and the completion of a speaking end-of-module assignment. Towards the end of the module, students receive an email offering them a choice of when to attend (date and time) their end-of-module speaking assignment, which is conducted in a synchronous online setting. Students who do not pick one of the proposed timeslots, are allocated to a default session (default allocation).

It was hypothesised that students who did not choose their own, convenient date for taking the final speaking assignment and received a default session are the most at-risk of not attending the compulsory end-of-module assignment and in consequence failing to complete the module. Those students were provided with a support intervention aiming to motivate participation to the end-of-module assignment and module completion. Only those students who had made sufficient progress in the continuous assessment component throughout the module, and therefore had a reasonable chance of passing, were included in the support interventions.

The action was conceived by the Language Studies programme Student Support Team Faculty Lead and implemented within available resources by the Student Support Team (SST). One attribute of the OU SST model is the qualification perspective and possibility of intervention across an entire qualification or subset of modules. For this action, lists of default allocation students from all four level 1 intermediate Languages modules identified for intervention were mixed and reordered by student number sequence, to a give a rough-and -ready randomisation. Based on the hypothesis that these students who have not chosen their date are the most at-risk of not attending, it was hoped to measure comparative success of the actions and recommend which was worth taking forward to another year.

Hypothesis 1: There will be no differences between intervention and control groups in terms of the end of the module attendance rates and performance.

Hypothesis 2: There will be no differences between intervention and control conditions in terms of passing the module.

Method

Interventions

Three support interventions were designed to test the aforementioned hypotheses. In the first intervention (Experimental Condition 1), students received an email message from the SST. SSTs monitor student progress and are an important point-of-call for students who may need additional support and guidance to complement that which they receive from their Associate Lecturer or 'tutor'. The email message sent from SST noted that a respective student had not actively selected the end of the module assessment slot, reminding the student to choose an end of module assessment date. Links to preparation materials on the relevant module website were also provided in the email, in a way comparable to the motivational emails by Inkelaar and Simpson (2015). In the second intervention (Experimental Condition 2), students received an outbound call from the Learner Support staff in the SST to check whether they were aware of the date and time of the end of the module assessment and prompt them to contact their tutor for any questions or additional support. Students who could not be reached, either due to no telephone number available or their voicemail could not be accessed, were excluded from the analysis (N=7). In the third intervention (Experimental Condition 3), an email was sent to the student's Associate Lecturer drawing their attention to the fact that certain named student/s had not picked their end-of-module session. Given that tutors are aware of their students' progress, they were left to decide whether they should contact students and in what ways, though optional message wording was provided. The control condition received no additional action and acted as the control group.

Ethical permissions

The OU Policy on Ethical use of Student Data for Learning Analytics (see http://bit.ly/10zbH54) states that OU students consent to the use of their data at the point of reservation or registration on to a module or qualification. Learning analytics data is only used to support students' learning and enhance the University's overall provision. Ethical advice was also gained by the university's ethics committee. Ethic experts proposed that the students' informed consent as stated in the policy was sufficient and that there was no need for further contact with students in relation to ethical approvals for implementing this study.

Sample

The sample of this study was 80 students who were given a "'default allocation" for attending the end-of-module assessment. Seven students in EC2 were not included in the analysis as they could not be reached through phone. The final sample size was 73 students. In the months preceding the end-of-module speaking assessment, the SST procured a list of default allocation students on each of the level 1 intermediate Languages modules, a pool of 308 at-risk students. To examine the impact of the support intervention on passing/failing the module the list was cross-checked with the students who at that point had submitted three of their four assignments throughout the module. This yielded a list of 80 students between the French, Spanish, German and Italian modules. The lists were mixed, put into unique student number order and allocated to one of the intervention or control conditions, that is to say, 20 students for each of the experimental condition and the control condition. In other words, these identified students were relatively successful and with the possibility of passing the module, but for some reason had not yet registered for the final speaking assessment. We hoped that a motivational action would encourage those students to take the final end-of-module speaking assessment and complete the module.

Data analysis

Descriptive and inferential statistics were used to examine the impact of the support intervention on students' end-of-module assessment and pass rates. In terms of gender, 47.9% (N = 35) were male and 52.1% (N = 38) were female. A great majority of students (43.8%) was in the age band 17 to 30 years old, 31.5% 31 to 45 years old, and the remaining of the sample (24.7%) 46 to 76 years old. Students taking the French module comprised 19.2% of the sample, the German module 39.7%, the Spanish module 26%, and the Italian module 15.1 %. A great majority of students completed the end of the module exam (N = 64, 87.7%) and managed to pass the modules (N = 60, 82.2%).

Results

Prior to performing any statistical comparisons and in order to confirm that the allocation of participants to one of the four conditions was indeed random, we ran a series of Chi-square tests to compare the four groups in terms of basic demographic characteristics. No significant differences were found in relation to whether a student was new to the OU or had already attended other online modules, age band, previous education, ethnic group and socio-economic status. Significant differences were found in relation to gender and number of students attending a specific module in each condition. In EC1, the majority of students were taking the French language module, in EC2 the Spanish language module and in the Control group the Italian language module. In terms of gender, more male students were found in EC1 and less in EC3. More female students were found in EC3-4. These differences were considered minor and less likely to have affected the randomization of participants in the four conditions.

A one-way ANOVA was performed to examine differences between conditions in relation to the end-of- module exam. Descriptive statistics point to a better performance of EC1 (M = 69.8, SD = 31) and 2 (M = 67.5, SD = 15.4) as opposed to EC3 (M = 55.1, SD = 32.7). The control group showed a similar mean value to EC1 and EC2 (M = 67.5, SD = 35.6). No significant effect of support intervention on students' end-of-module performance was found ($F (3, 68) = 0.89$, $p = .45$, NS). A chi-square test of independence was performed to examine whether the support intervention had an impact on the number of students attending the end-of-module examination. The frequency distribution indicates that two students in EC1, four students in EC3 and three students in the control condition did not attend the examination. No statistically significant differences were found between the four conditions (χ^2 (3, N = 73) = 3.1, $p = .37$, NS). A chi-square test of independence was also performed to evaluate differences between conditions in passing the module. No relationships were found between the four conditions and pass/failure conditions (χ^2 (3, N = 73) = 5.0, $p = .17$, NS).

The outcomes of this analysis revealed no significant differences between the three intervention types and control condition in relation to support provided to at-risk students before the end-of-module assessment. Given the higher end of the module mean scores, it could be argued that the first two interventions (email and outbound calls) were more effective than the third intervention (email sent to tutors). Yet, none of the three types of intervention was found to have an impact on the end-of-module assessment attendance, performance and module completion when compared to the control condition.

➔ Insert Table 1 about here

Discussion

Although considerable progress has been observed in predicting learners who are at-risk, (Calvert, 2014; Macfadyen & Dawson, 2010; Tempelaar et al., 2015), it is still unclear what type of interventions should be designed and put into practice to effectively support different groups of distance learners. Building on two studies in this journal (Inkelaar & Simpson, 2015; Woods & Keeler, 2001), this short-scale study described an example of how evidence-based research, in particular Randomised Control Trials (RCTs), might be used in open and distance education to effectively test the impact of various experimental conditions relative to the control condition.

Our small scale intervention indicated no significant differences between the three experimental conditions and the control condition. Before one can conclude that targeted emails, phone calling, or RCTs are a waste of money, we suggest three reasons why conducting this study was actually very useful for evidence-based research. First of all, one of the reasons for the lack of impact might be related to the short time duration and the relatively small intervention. Perhaps students had already made up their minds (or not) to attend the final exam, whereby a small nudge might have been insufficient to change their intentions. Alternatively, perhaps the nudge might have strengthened the self-belief of at-risk students not to register for the end of module assessment. Finally, without doing an RCT the automatic assumption of some stakeholders across HE institutions might be that an email or phone call always have a positive or at least neutral effect on students' intentions. Our study indicates that careful consideration is needed for whom, how, for how long, and how intense support is provided to make an impact.

The fact that students were randomly assigned to each condition eliminated bias and offset potential Hawthorne or John Henry effects. Yet, it is not known whether students had contact and shared their study experiences in out of the university spaces, such as social media, and thus became aware of being in different conditions. Such conditions would more likely bias the study findings as students might have altered their behaviour knowing that they are observed or belonged in the control group. However, these interactions are out of the research team's control and could only be waived by recruiting a large sample size of students in each condition. This could eliminate potential impact of some students being aware that they were treated differently.

Our study would benefit from a qualitative understanding of the reasons why no significant differences were observed by, for example, interviewing students from each condition. Such an account would illuminate the kind of interventions students perceive as useful and supportive of their learning. Towards this direction, we have set up a large-scale evaluation to examine the impact of specific intervention strategies on students' retention with teachers from 25 modules. This project will make use of experimental methodologies including A/B testing and RCTs to evaluate the use of predictive data by teachers and couple this with qualitative data (e.g., interviews) to understand why certain interventions might work while others may not.

Conclusions

For some academics implementing a small-scale RCT in practice might be challenging while for others it might be relatively straightforward. For academics familiar with educational research and who have sufficient data interpretation skills, the application of an RCT to practice will more likely be relatively straightforward to implement using principles of quantitative research. For academics who do not have these skills, collaboration in evaluating outcomes from an RCT with colleagues might help them to make informed suggestions for follow-up interventions. The use of evidence-based research, in particular RCTs, in this article aims to encourage academics to apply such approaches to their practice to help at-risk students by showcasing how an RCT could be designed, conducted and analysed. These academics might lead the way and inspire other colleagues to integrate this methodology to their practice, generate enthusiasm and tackle resistance.

Academics or other individuals interested in applying RCTs to their practice need to be knowledgeable of the strengths and weaknesses of this approach and consider these especially when analysing and interpreting their study's outcomes. This small-scale study aims to bring to the fore pros and cons of the use of RCTs in open learning and spark discussions around the use of RCTs as a robust method of evaluating the effectiveness of online learning analytics interventions in HE. In contrast to face-to-face educational RCT evaluation, RCTs might be particularly beneficial in web-based implementations, overcoming some of their inherent limitations. Due to the lack of face-to-face interactions, students who study online are less likely to become aware of the condition they are in, while researchers can have better control over each condition and its characteristics. According to Norman (2010), RCTs are useful in education when examining standardized interventions such as simulation in clinical conditions and web-based learning (Norman, 2010). The web-based implementation of learning allows better manipulation of variables, including keeping variables that are not under study constant. For example, online feedback could be provided to a specific VLE learning activity as opposed to the same activity without the addition of feedback. Two groups of randomly allocated students then interact with the activities for the same period of time and their interactions are monitored. Researchers could then confidently conclude on whether what is observed is due to the intervention or any other confounding variables, as all learning conditions - apart from the addition of feedback - are kept constant for both groups while potential impact of the learners' characteristics on the intervention is waived by the random allocation of learners. Randomisation of participants can largely remove the risk of bias in any study's findings by evenly allocating participants' characteristics across groups. RCTs can be used to showcase to policymakers which interventions are most or more effective than others for adoption and take up.

Despite the benefits from implementing RCT testing, the so-called “gold-standard” in research in education (Slavin, 2002; 2008; Torgerson & Torgerson, 2008), it is noted that the actual implementation might be complex due to organisational, technical, and perhaps ethical barriers. One such barrier might be the random assignment of learners to two or more groups. VLE systems might not allow access to teachers to perform such allocation and trace changes in learners' behaviour. Assuming that convenient conditions are in place, there might still be challenges related to the quality of the sample under study. Small samples (<200) might require additional tests to ensure equal distribution of learners' characteristics across conditions. In other cases, the process of gaining ethical permission to run an RCT might be complex and time-consuming and might even inhibit the research implementation if considered unethical. It is often suggested that participants in the control group might be disadvantaged as opposed to the intervention group by not receiving the proposed intervention. Yet, without knowing whether the impact of an intervention is positive or not such arguments cannot hold. The potential for robust and reliable evidence in relation to the RCT evaluation of distance learning interventions (Slavin, 2008) might well outperform these challenges.

Also, the use of RCTs in education, as opposed to healthcare, has been contested due to a lack of providing insights as to why positive impact is observed or not. Deploying additional methodologies is critical in this respect as they can explain the outcomes of RCTs, especially in cases when an RCT failed to find evidence of impact. Qualitative approaches such as interviews and questionnaires could clarify why, for example, a positive effect has occurred at a given time and setting and what the mechanisms are that underpin how it works.

Such information would be useful when repeating an intervention elsewhere (Cartwright & Hardie, 2012).

In conclusion, drawing from our own practical experience and the development of the Analytics4Action Evaluation Framework (A4AEF) (see section: RCTs in Education), the implementation of RCTs in distance learning institutions seems to require both strategic support from senior management and trust of academics and teachers. Although this can be a challenging and demanding endeavour, it could be achieved through constant communication and collaboration with key stakeholders (e.g., weekly or monthly meetings, email communication) to provide real-time support when needed and safeguard the effective implementation and evaluation of RCTs or other evidence-based research approaches in education. In respect to this study, we shared our approach and findings with stakeholders across the university in a workshop. Involved stakeholders embraced positively the use of experimental methodology to evaluate web-based learning analytics interventions. Yet, they noted that such studies should be well-designed, with clearly defined aims and objectives, tackle specific learning and teaching challenges, and benefit the university's performance.

References

Aboalshamat, K., Hou, X.-Y., & Strodl, E. (2015). The impact of a self-development coaching programme on medical and dental students' psychological health and academic performance: a randomised controlled trial. *BMC Medical Education, 15*(1), 1-13. doi: 10.1186/s12909-015-0412-4

Arbaugh, J. B. (2014). System, scholar, or students? Which most influences online MBA course effectiveness? *Journal of Computer Assisted Learning, 30*(4), 349-362. doi: 10.1111/jcal.12048

Rienties, B., Boroowa, A., Cross, S., Kubiak, C., Mayles, K., & Murphy, S. (2016). Analytics4Action Evaluation Framework: a review of evidence-based learning analytics interventions at Open University UK. *Journal of Interactive Media in Education, 1*(2), 1-12. doi: 10.5334/jime.394

Rienties, B., Cross, S., & Zdrahal, Z. (2016). Implementing a Learning Analytics Intervention and Evaluation Framework: what works? In B. K. Daniel (Ed.), *Big Data and Learning Analytics in Higher Education: Current Theory and Practice* (pp. 147-166). Heidelberg: Springer.

Rienties, B., Giesbers, S., Lygo-Baker, S., Ma, S., & Rees, R. (2016). Why some teachers easily learn to use a new Virtual Learning Environment: a Technology Acceptance perspective. *Interactive Learning Environments, 24*(3), 539-552. doi: 10.1080/10494820.2014.881394

Rienties, B., & Toetenel, L. (2016). The impact of learning design on student behaviour, satisfaction and performance: a cross-institutional comparison across 151 modules. *Computers in Human Behavior, 60*, 333-341. doi: 10.1016/j.chb.2016.02.074

Tempelaar, D. T., Rienties, B., & Giesbers, B. (2015). Stability and sensitivity of Learning Analytics based prediction models*Proceedings of 7th International conference on Computer Supported Education* (pp. 156–166). Lisbon: CSEDU.

Baker, R. S. (2010). Data mining for education. *International encyclopedia of education*, *7*, 112-118.

Bienkowski, M., Feng, M., & Means, B. (2012). Enhancing teaching and learning through educational data mining and learning analytics: An issue brief. *US Department of Education, Office of Educational Technology*, 1-57.

Calvert, C. E. (2014). Developing a model and applications for probabilities of student success: a case study of predictive analytics. *Open Learning: The Journal of Open, Distance and e-Learning*, *29*(2), 160-173. doi: 10.1080/02680513.2014.931805

Cartwright, N. & Hardie, J. (2012). Evidence-based policy a practical guide to doing it better. USA: Oxford University Press.

Christie, H., Munro, M., & Fisher, T. (2004). Leaving university early: exploring the differences between continuing and non-continuing students. *Studies in Higher Education*, *29*, 617-636.

Ferguson, R., & Buckingham Shum, S. (2012). *Social learning analytics: five approaches*. Paper presented at the Proceedings of the 2nd International Conference on Learning Analytics and Knowledge, Vancouver, British Columbia, Canada.

Franssen, R., & Nijhuis, J. (2011). Exploring Student Attrition in Problem-Based Learning: Tutor and Student Perceptions on Student Progress. In P. Van den Bossche, W. H. Gijselaers, & R. G. Milter (Eds.), *Building Learning Experiences in a Changing World* (Vol. 3, pp. 139-146): Springer Netherlands.

González-Torres, A., García-Peñalvo, F. J., & Therón, R. (2013). Human–computer interaction in evolutionary visual software analytics. *Computers in Human Behavior, 29*(2), 486-495. doi: 10.1016/j.chb.2012.01.013

Hanson, J. (2009). Displaced but not replaced: the impact of e-learning on academic identities in higher education. *Teaching in Higher Education, 14*(5), 553-564. doi: 10.1080/13562510903186774

Harford, T. (2014, 25 April 2014). The random risks of randomised trials. *The Financial Times Magazine*. Retrieved from http://www.ft.com/cms/s/2/59bb202c-ca7b-11e3-8a31-00144feabdc0.html#axzz2zvfIZC7h

Hattie, J. (2009). *Visible Learning: A synthesis of over 800 meta-analyses relating to achievement.* New York: Routledge.

Hess, F. M., & Saxberg, B. (2013). *Breakthrough Leadership in the Digital Age: Using Learning Science to Reboot Schooling*. Thousand Oaks: Corwin Press.

Hickey, D. T., Kelley, T. A., & Shen, X. (2014). *Small to big before massive: scaling up participatory learning analytics.* Paper presented at the Proceedings of the Fourth International Conference on Learning Analytics And Knowledge.

Hommes, J., Arah, O. A., de Grave, W., Bos, G., Schuwirth, L., & Scherpbier, A. (2014). Medical students perceive better group learning processes when large classes are made to seem small. *PLOS One, 9*(4), e93328. doi: 10.1371/journal.pone.0093328

IET Student Statistics and Survey Team. (2014). What reasons do students give for not completing modules? Non-returners Research Findings and Outcomes.

Ilic, D., & Maloney, S. (2014). Methods of teaching medical trainees evidence-based medicine: a systematic review. *Medical Education, 48*(2), 124-135. doi: 10.1111/medu.12288

Inkelaar, T., & Simpson, O. (2015). Challenging the 'distance education deficit' through 'motivational emails'. *Open Learning: The Journal of Open, Distance and e-Learning, 30*(2), 152-163. doi: 10.1080/02680513.2015.1055718

Macfadyen, L. P., & Dawson, S. (2010). Mining LMS data to develop an "early warning system" for educators: A proof of concept. *Computers & Education, 54*(2), 588-599. doi: 10.1016/j.compedu.2009.09.008

MacNeill, S., Campbell, L. M., & Hawksey, M. (2014). Analytics for Education. *Journal of Interactive Media in Education*. doi: 10.5334/2014-07

McMillan, J. H., & Schumacher, S. (2014). *Research in education: Evidence-based inquiry*. Harlow: Pearson Higher Ed.

Norman, G. (2010). Is experimental research passé? *Advances in Health Sciences Education: Theory and Practice, 15* (3), 297–301.

Papamitsiou, Z., & Economides, A. (2014). Learning Analytics and Educational Data Mining in Practice: A Systematic Literature Review of Empirical Evidence. *Educational Technology & Society, 17*(4), 49–64.

Piderit, S. K. (2000). Rethinking Resistance and Recognizing Ambivalence: A Multidimensional View of Attitudes toward an Organizational Change. *The Academy of Management Review, 25*(4), 783-794. doi: 10.2307/259206

Richins, M. L. (1983). Negative Word-of-Mouth by Dissatisfied Consumers: A Pilot Study. *Journal of Marketing, 47*(1), 68-78.

Simpson, O. (2004). The impact on retention of interventions to support distance learning students. *Open Learning: The Journal of Open, Distance and e-Learning, 19*(1), 79-95. doi: 10.1080/0268051042000177863

Simpson, O. (2013). *Supporting students for success in online and distance education*. New York, NY: Routledge.

Siroker, D., & Koomen, P. (2013). *A/B Testing: The Most Powerful Way to Turn Clicks Into Customers*. New Jersey: John Wiley & Sons.

Slavin, R. E. (2002). Evidence-Based Education Policies: Transforming Educational Practice and Research. *Educational Researcher, 31*(7), 15-21. doi: 10.2307/3594400

Slavin, R. E. (2008). Perspectives on evidence-based research in education—What works? Issues in synthesizing educational program evaluations. *Educational Researcher, 37*(1), 5-14. doi: 10.3102/0013189X08314117

Tempelaar, D. T., Rienties, B., & Giesbers, B. (2015). In search for the most informative data for feedback generation: Learning Analytics in a data-rich context. *Computers in Human Behavior, 47*, 157-167. doi: 10.1016/j.chb.2014.05.038

Tobarra, L., Robles-Gómez, A., Ros, S., Hernández, R., & Caminero, A. C. (2014). Analyzing the students' behavior and relevant topics in virtual learning communities. *Computers in Human Behavior, 31*(0), 659-669. doi: 10.1016/j.chb.2013.10.001

Torgerson, D. J., & Torgerson, C. (2008). *Designing randomised trials in health, education and the social sciences: an introduction*. London: Palgrave Macmillan.

Wetzer, I. M., Zeelenberg, M., & Pieters, R. (2007). “Never eat in that restaurant, I did!”: Exploring why people engage in negative word-of-mouth communication. *Psychology & Marketing*, *24*(8), 661-680.
Woods, R., & Keeler, J. (2001). The Effect of Instructor's Use of Audio E-mail Messages on Student Participation in and Perceptions of Online Learning: A preliminary case study. *Open Learning: The Journal of Open, Distance and e-Learning*, *16*(3), 263-278. doi: 10.1080/02680510120084977
Ziliak, S., & Teather-Posadas, E. R. (2014). The Unprincipled Randomization Principle in Economics and Medicine. In G. DeMartino & D. N. McCloskey (Eds.), *Oxford University Press Handbook on Professional Economic Ethics* New York: Oxford.

Table 1 *Mean scores in 4 TMAs, final exam, and percentage of students passing the module.*

	TMA1	TMA2	TMA3	TMA4	Final exam	# passed	% passed
Condition 1	79.3	77.0	79.2	72.8	69.8	17/20	85.0
Condition 2	79.1	70.5	75.3	72.4	67.5	13/13	100.0
Condition 3	82.9	77.2	75.0	74.7	55.1	14/20	70.0
Condition 4	80.5	76.6	78.0	77.2	67.5	16/20	80.0

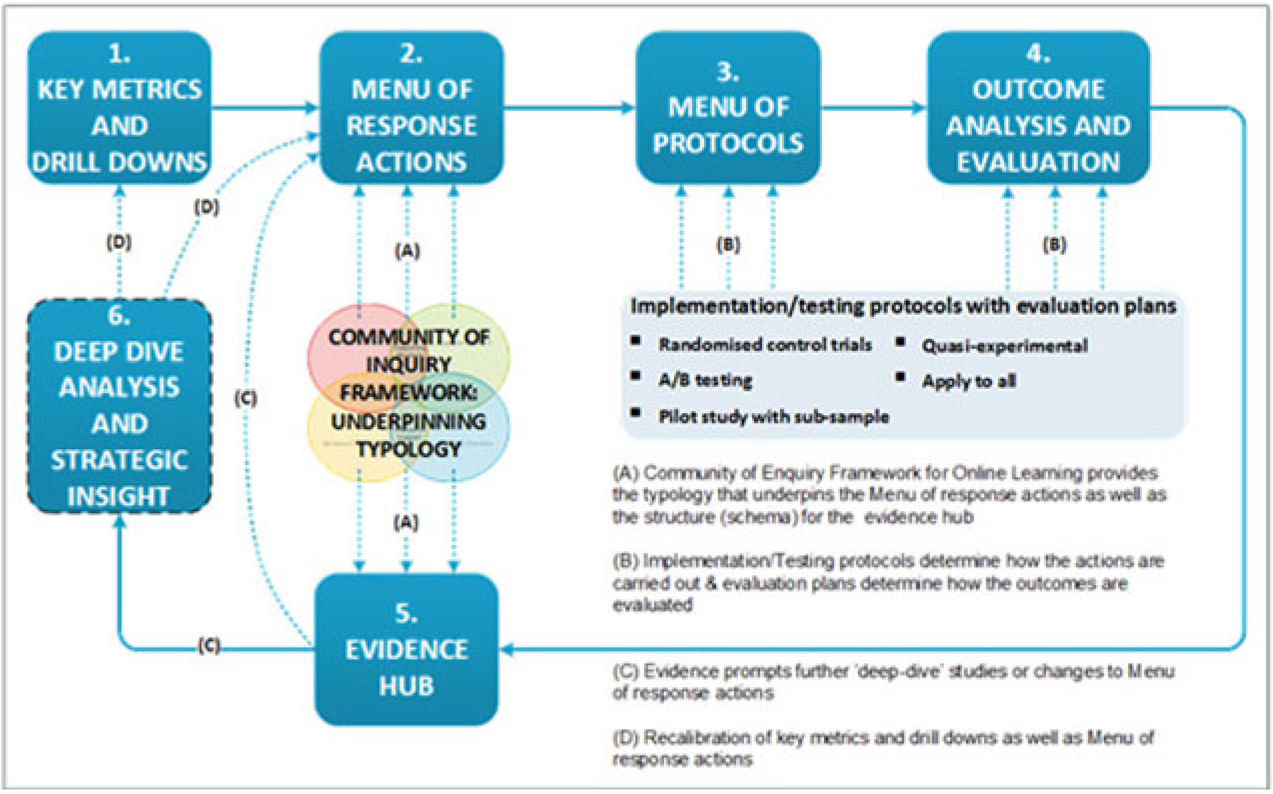

Figure 1 Analytics4Action Evaluation Framework

Source: Rienties, Cross, & Zdrahal, 2016.