Implementing a Learning Analytics Intervention and Evaluation Framework: what works?

Bart Rienties, Simon Cross, Zdenek Zdrahal. The Open University UK, bart.rienties@open.ac.uk.

Publication forthcoming in:

Rienties, B., & Cross, S., Zdrahal, Z. (Forthcoming February 2016). Implementing a Learning Analytics Intervention and Evaluation Framework: what works? In Motidyang, B., & Butson, R. (Eds.) *Big Data and Learning Analytics in Higher Education Current Theory and Practice*. Springer, Heidelberg.

Abstract

Substantial progress in learning analytics research has been made in recent years to predict which groups of learners are at-risk. In this chapter we argue that the largest challenge for learning analytics research and practice still lies ahead of us: using learning analytics modelling, which types of interventions have a positive impact on learners' Attitudes, Behaviour and Cognition (ABC). Two embedded case-studies in social science and science are discussed, whereby notions of evidence-based research are illustrated by scenarios (quasi-experimental, A/B-testing, RCT) to evaluate the impact of interventions. Finally, we discuss how a Learning Analytics Intervention and Evaluation Framework (LA-IEF) is currently being implemented at the Open University UK using principles of design-based research and evidence-based research.

Introduction

Many institutions and organisations across the globe seem to have high hopes that analytics and big data can make their organisations fit-for-purpose, flexible, and innovative. Learning analytics applications in education are expected to provide institutions with opportunities to support learner progression and to enable personalised, rich learning on a large scale (Bienkowski, Feng, & Means, 2012; Hickey, Kelley, & Shen, 2014; Siemens, Dawson, & Lynch, 2013; Tempelaar, Rienties, & Giesbers, 2015; Tobarra, Robles-Gómez, Ros, Hernández, & Caminero, 2014). With the increased availability of large datasets, powerful analytics engines (Tobarra et al., 2014), and skilfully designed visualisations of analytics results (González-Torres, García-Peñalvo, & Therón, 2013), institutions may be able to use the experience of the past to create supportive, insightful models of primary (and perhaps real-time) learning processes (Baker, 2010; Ferguson & Buckingham Shum, 2012; Papamitsiou & Economides, 2014; Stiles, 2012). According to Bienkowski et al. (2012, p. 5), "education is getting very close to a time when personalisation will become commonplace in learning", although several researchers (García-Peñalvo, Conde, Alier, & Casany, 2011; Greller & Drachsler, 2012; Stiles, 2012; Tempelaar et al., 2015) indicate that most institutions may not be ready to exploit the variety of available datasets for learning and teaching, or have staff with the required skills in learning design.

Nevertheless, substantial progress in learning analytics research has been made in the last two-three years in using a range of advanced computational techniques (e.g., predictive modelling, machine learning, Bayesian modelling, social network analysis, cluster analysis) to predict which students are going to pass a course, and which (groups of) learners are at-risk (Agudo-Peregrina, Iglesias-Pradas, Conde-González, & Hernández-García, 2014; Aguiar, Chawla, Brockman, Ambrose, & Goodrich, 2014; Calvert, 2014; Gasevic, Zouaq, & Janzen, 2013; Tempelaar et al., 2015; Tobarra et al., 2014; Wolff, Zdrahal, Herrmannova, Kuzilek, & Hlosta, 2014). What these studies have in common is that by collecting (longitudinal) data from a range of sources the accuracy in predicting learning performance is increasing with every new piece of information about the learner and his/her learning.

In this chapter we argue that the largest challenge for learning analytics research and practice still lies ahead of us: *which types of interventions have a positive impact on learners' Attitudes, Behaviour and Cognition (ABC) using learning analytics modelling?* Most of the current literature seems to be focussed on testing and applying principles of learning analytics using convenience sampling, whereby a particular context or sample of students in a course is taken to illustrate the predictive power of learning analytics approaches. However, most research seems to lack a robust *Design-Based Research* (Collins, Joseph, & Bielaczyc, 2004; Rienties & Townsend, 2012) or *evidence-based approach* (e.g., Randomised Control Trials, A/B testing, pre-post retention modelling) of testing and validating claims and arguments (e.g., Hess & Saxberg, 2013; McMillan & Schumacher, 2014; Rienties et al., 2012; Rienties, Giesbers, Lygo-Baker, Ma, & Rees, 2014; Slavin, 2008). Indeed, according to Collins et al. (2004, p. 21), design experiments should feature and "bring together two critical pieces in order to guide us to better educational refinement: a design focus and assessment of critical design elements".

Although we acknowledge that learning analytics can provide a powerful learning experience for different groups of learners and purposes, in this chapter we primarily focus on the use of learning analytics for students-at-risk. If organisations are going to adopt learning analytics approaches in order to improve learning, then the research community needs to provide evidence-based results that highlight that learning analytics approach can: 1) identify learners at-risk; 2) deliver (personalised) intervention suggestions that work; 3) be cost-effective. There is an urgent need to develop an evidence-based framework of learning analytics that can help to inform students, researchers, educators, and policy makers about which types of interventions work well under which conditions, and which do not. In this chapter, we will work towards establishing principles of a *Learning Analytics Intervention and Evaluation Framework* (LA-IEF) that will be tested and validated at the largest university in Europe (in terms of enrolled learners), namely the Open University UK (Calvert, 2014; Richardson, 2012a). Firstly, we will provide a short literature review of contemporary learning analytics studies, followed by an argument in support of needing more robust evidence-based research. Secondly, we will provide two case-studies how principles of evidence-based research could be implemented in learning analytics. These two case-studies will provide stepping stones how to build an LA-IEF model, adjust and apply the framework into their own practice.

Attitudes, Behaviour, Cognition (ABC model) and Learning Analytics

Given that learning analytics is a relatively new research field using a range of interdisciplinary perspectives that did not exist before 2010, it is not surprising that most of the research efforts have thus far focussed on identifying and raising awareness of the conceptions, boundaries, and generic approaches of learning analytics (Ferguson, 2012; Papamitsiou & Economides, 2014; Wise, 2014). Alternatively, we use a simple attitudes-behaviour-cognition (ABC) model to conceptualise the possible impacts of learning analytics on learning. *Attitudes* of learners can have positive impacts on behaviour (Giesbers, Rienties, Tempelaar, & Gijselaers, 2013; Jindal-Snape & Rienties, 2016; Pintrich & De Groot, 1990; Rienties & Alden Rivers, 2014; Rienties et al., 2012; Tempelaar et al., 2015), such as intrinsic motivation, self-efficacy, curiosity, or goal-orientation, which can trigger learners into action. In contrast, negative attitudes may hamper learners (Martin, 2007; Pekrun, Goetz, Frenzel, Barchfeld, & Perry, 2011; Rienties & Alden Rivers, 2014; Tempelaar, Niculescu, Rienties, Giesbers, & Gijselaers, 2012), A-motivation, boredom, anxiety, or stress, can all lead to a restriction of action or reduction in engagement, or even withdrawal from learning. For example, in a study of 730 students Tempelaar et al. (2012) found that positive learning emotions contributed favourably to becoming an intensive online learner, while negative learning emotions, like boredom, contributed adversely to learning behaviour. Similarly, in an online community of practice of 133 instructors supporting EdD students, Nistor et al. (2014) found that self-efficacy (and expertise) of instructors predicted online contributions.

Many learning analytics applications use *behaviour* data generated from learner activities, such as the number of clicks (Siemens, 2013; Wolff, Zdrahal, Nikolov, & Pantucek, 2013), learner participation in discussion forums (Agudo-Peregrina et al., 2014; Macfadyen & Dawson, 2010), or (continuous) computer-assisted formative assessments (Papamitsiou, Terzis, & Economides, 2014; Tempelaar et al., 2012; Tempelaar et al., 2015; Wolff et al., 2013). User behaviour data are frequently supplemented with background data retrieved from Virtual Learning Environment (VLE) (Macfadyen & Dawson, 2010) and other learner admission systems, such as accounts of prior education (Arbaugh, 2014; Calvert, 2014; Richardson, 2012a). For example, in one of the first learning analytics studies focused on 118 biology students, Macfadyen and Dawson (2010) found that some (e.g., # of discussion messages posted, # assessments finished, # mail messages sent) VLE variables but not all (e.g., time spent in the VLE) were useful predictors of learner retention and academic performance.

However, a recent special issue on Learning Analytics in *Computers in Human Behaviour* (Conde & Hernández-García, 2015) indicates that simple learning analytics metrics (e.g., # of clicks, # of downloads) may actually hamper learning analytics research. For example, using a longitudinal data analysis of 120+ variables from three different VLE systems and a range of motivational, emotions and learning styles indicators, Tempelaar et al. (2015) found that most "simple" VLE learning analytics metrics provided limited insights into the complex learning dynamics over time. In contrast, learning motivations and emotions (attitudes) and activities done by learners during continuous assessments (behaviour) provided an opportunity for teachers to help at-risk learners at a relatively early stage of their learning journey. Similarly, in a more fine-grained study Giesbers et al.

(2013) found that discussion forum and synchronous videoconference behaviour of 110 students, motivation of students (attitudes) and tool usage (behaviour) significantly influenced cognition.

Finally, *cognition* is conceptualised as a wide construct, such as learning a new skill, understanding and evaluating a theoretical concept, or applying something learned into practice. In Western education, cognition is often translated into performance on summative assessments (Agudo-Peregrina et al., 2014; Calvert, 2014; Macfadyen & Dawson, 2010; Tempelaar et al., 2015), such as exams, quizzes, multiple-choice tests, or open essays. However, “evidence” of cognition can also be found in more formative learning activities (Ferguson & Buckingham Shum, 2012; Knight, Buckingham Shum, & Littleton, 2013; Rienties et al., 2012), such as discussion forum postings, articulations of discourse in web-videoconferences, blog postings, or reflections. In other words, depending on the conceptualisation of cognition, learning analytics models may take a relatively narrow approach (i.e., did a learner pass a module) or a broader approach (e.g., is there evidence of critical evaluation skills in a string of discussion forum messages on topic X).

A need for evidence-based research in learning analytics

While these initial research papers have substantially contributed to forming a foundation of learning analytics as a new research field, the proof of the pudding will be whether learning analytics is able to make a fundamental, measurable impact on attitudes, behaviour, and cognition of learners, teachers and institutions in general. To the best of our knowledge, no learning analytics studies are available that can be described as *evidence-based research* (MacNeill, Campbell, & Hawksey, 2014; Slavin, 2002, 2008). A basic approach of evidence-based research in education is to use scientific methods based upon experiments to apply the best available, significant and reliable, evidence to inform educational decision making. While in many fields like medicine, agriculture, transportation, and technology, the process of development, rigorous evaluation, and sharing of results to practice using randomised experiments (Torgerson & Torgerson, 2008) and A/B testing (Siroker & Koomen, 2013) have provided an unprecedented innovation in the last fifty years (Slavin, 2002), educational research and learning analytics in particular has not adopted evidence-based research principles en masse (Hess & Saxberg, 2013; McMillan & Schumacher, 2014; Torgerson & Torgerson, 2008).

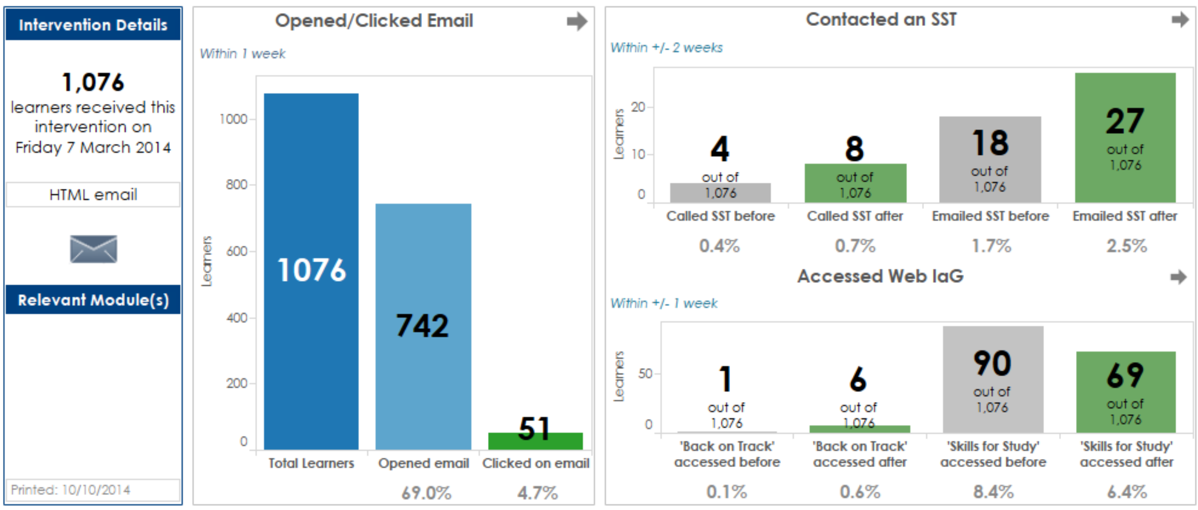
A major potential problem of descriptive or correlational studies is the ability to generalise the findings beyond the respective context in which particular learning analytics study has been conducted (Arbaugh, 2005, 2014; Hattie, 2009; Rienties, Toetenel, & Bryan, 2015). Especially the issue of selection bias is a concern in early learning analytics studies. A recent meta-analysis of 35 empirical studies (Papamitsiou & Economides, 2014) indicated that most learning analytics studies have focussed on analysis of single-module or single discipline, using contexts of teachers and/or learners who are keen to share their practice for research. Although these studies provide relevant initial insights of principles of data analysis techniques, and testing the proof of concepts of data visualisation tools, without random assignment of learners (and teachers) into two or more conditions it is rather difficult to provide evidence of impact (McMillan & Schumacher, 2014; Slavin, 2008; Torgerson & Torgerson, 2008). For example, an engaged and enthusiastic teacher who used a data visualisation tool to provide feedback to her learners might trigger a direct response from them to follow-up on the suggested feedback. However, replicating the study with a different group of learners, a different teacher, or even a different context might lead to completely different results. Similarly, attributing a positive learning effect of a particular data visualisation tool in a pre-post-test design without a clear A-B testing or RCT design to compare the data visualisation tool with another tool or approach might lead some researchers to conclude a positive effect of this tool, although other elements in the learning design could have had a contribution to this positive effect.

In the second part of this chapter, we will focus on two embedded case-studies in two introductory courses in social science and science. The first case-study was selected as an example of how institutions can use learning analytics approaches to send early-warning emails to students potentially at risk (Inkelaar & Simpson, 2015). The second case-study was selected to illustrate how advanced in-house learning analytics applications can be used to identify which students are at-risk when multiple sources of static and dynamic data are available. Currently over 200,000 learners study at the Open University UK, primarily using distance education formats. An embedded case-study approach is undertaken to examine the characteristics of a single individual unit (recognising its individuality and uniqueness), namely, a learner, a group, or an organisation. Yin (2009) emphasised that a case-study investigates a phenomenon in-depth and in its natural context. Our two case-studies help us unpack and understand the challenges associated with evaluating interventions (which feeds in to the LA-IEF we developed). Using concepts of evidence-based research we will propose several scenarios how researchers could get a deeper understanding of impact of these interventions.

Case-study 1 Sending emails to learners at-risk

In an introductory social science module lasting 36 weeks, 1076 learners were identified by their educational profiles (e.g., low prior education, low assessment scores on previous modules) and engagement in the course (e.g., # VLE clicks in the last two weeks) as potentially "at-risk". This group received an email after four weeks with the intention "to encourage reflection on study progress as a mid-module progress check. Learners signposted to web resources such as "back on track". Following this intervention, if learners were concerned about their progress, they could obtain further information have to move ahead, whom to contact, or signposted to the deferrals, withdrawals and cancelation website. Another (indirect) aim of this intervention was to be able to link and measure engagements of learners through various ICT and administrative systems together to provide a more holistic perspective of the learners' journey throughout the module.

Figure 1 Impact of Email intervention on follow-up help-seeking behaviour



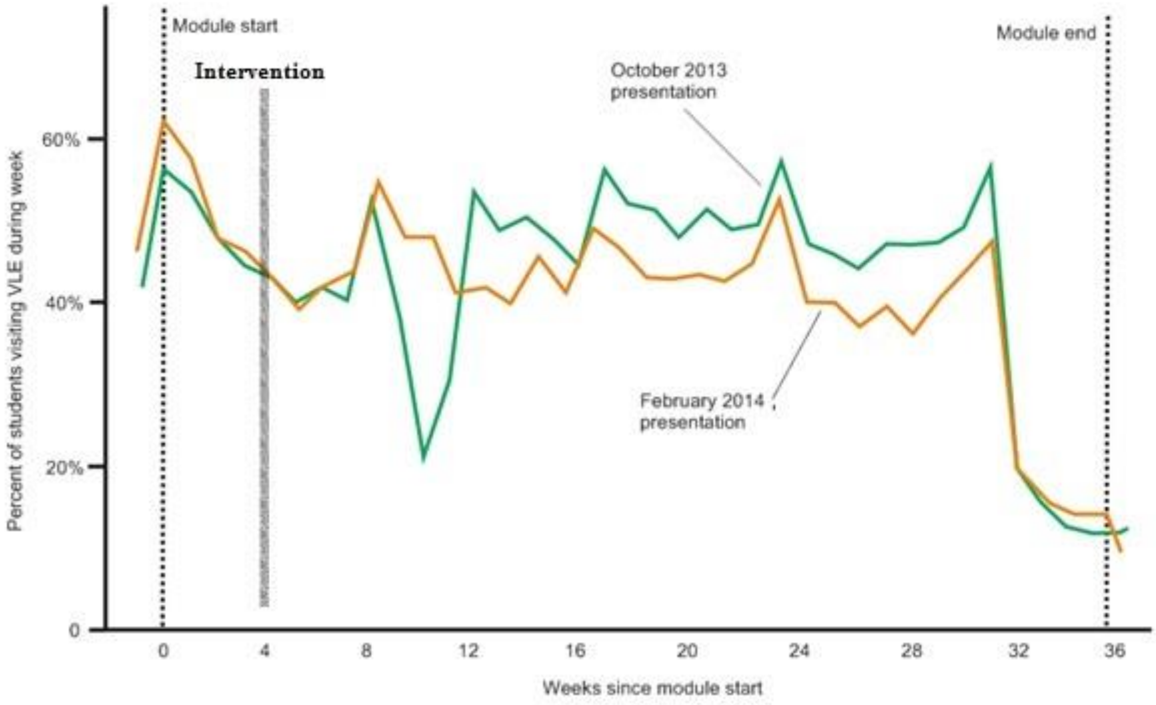
As indicated in Figure 1, of the 1076 learners who received the email, 742 (69%) learners opened the email within one week after receiving the email, while 51 (4.7%) learners followed up by clicking on one or more of the various links to further information (i.e., behaviour), such as the back-on-track website, skills for study website, and a contact for the Student Support Team (SST). In terms of back-on-track website, one learner accessed this website before the intervention, and six learners accessed this website within one week after the intervention. 90 learners accessed the study skills website before the intervention, and 69 learners accessed this website after the intervention. In terms of contact with SST, four learners called them before the intervention, while 8 called within two weeks after the intervention. Similarly, 18 learners emailed SST before the intervention, and 27 mailed after the intervention.

Although this intervention at first glance may indicate a positive impact on expected behaviour, because more learners engaged (behaviour) with one of the two support websites and/or contacted the SST team in comparison to before the intervention. One has to be careful to conclude that this intervention "worked" in terms of attitudes, behaviour and/or cognition. First of all, the number of learners who followed up with the "expected" behaviour was relatively small (<5%), and whether this would be statistically significant and meaningful might be questioned. Second, it is plausible be that of the eight learners who called the SST, seven already planned to do this, or perhaps 25 learners had planned to call student support but after receiving the intervention decided that continuing with the module did not make sense and dropped out. Third, in line with self-selection problems in non-randomised interventions, perhaps those learners with positive attitudes (e.g., intrinsically motivated) were more inclined to follow-up with the suggested action (Richardson, 2012b; Rienties et al., 2012), while those learners who actually needed help might have ignored the message. In other words, researchers need to be transparent about the relationships and fit between expected impact (i.e., planned during the intervention design) on ABC and the sensitivity of the measure(s) being used. In order to mitigate some of these issues, in an evidence-based approach we propose five possible scenarios to provide evidence of impact of case-study 1, each with its own strengths and limitations.

Scenario 1 Comparison with previous implementation

A natural option to compare the impact of an intervention is to contrast the ABC of learners from a previous implementation of the module. As illustrated in Figure 2, the module implementation starting in February 2014 is compared to the previous implementation in October 2013. Looking only at the two weeks after the intervention was initiated in week 4, no substantial difference in engagement can be discerned between the two cohorts in terms of aggregate VLE engagement. Whether or not the intervention in week 4 worked (or not) depends on what kinds of relations we are looking for, and how these are measured. Furthermore whether these two cohorts were similar at the start could be questioned as the percentage of learners visiting the VLE during the respective week was higher before the intervention for the 2014 implementation, and seemed to follow a similar trend until four weeks after the intervention. The substantial dip of activity in the October 2013 after 10 weeks is probably due to the Christmas break, but from week 12 onwards learners in the October 2013 presentation had substantially more engagement in the VLE than those in the 2014 implementation. Perhaps these aggregate data visualisations may under- or overestimate the complex, dynamic underlying engagements of learners with different ABC. In terms of cognition, similar passing rates were achieved in both modules. In other words, comparing the effectiveness of an intervention with a previous implementation might seem appealing, but addressing cause and effect is likely to be difficult (Hess & Saxberg, 2013; Torgerson & Torgerson, 2008).

Figure 2 VLE engagement before and after intervention (2013 vs. 2014 implementation).



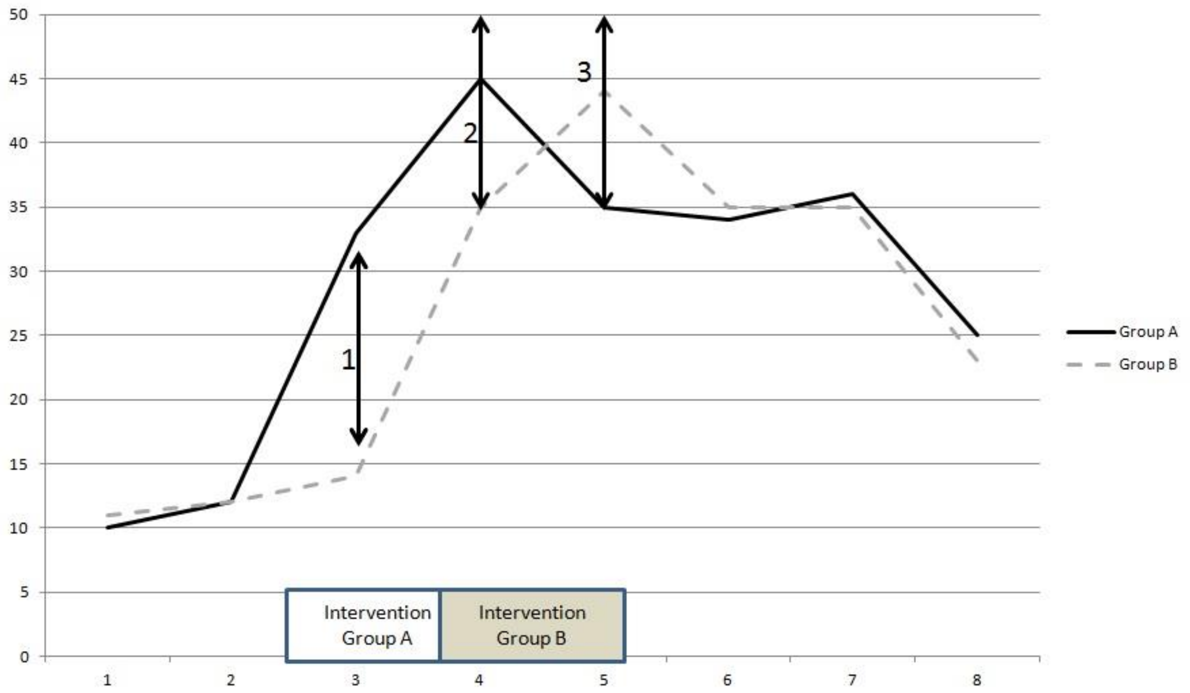
Scenario 2 Quasi-experimental follow-up

Institutions may not be able to conduct A/B testing or randomised control trials, due to potential ethical concerns (e.g., giving Group A a favourable treatment, while giving Group B a slightly less attractive treatment). When conducted well with appropriate control of confounding variables, quasi-experimental research could alleviate some of these issues as all learners in their respective cohort get the same treatment (Collins et al., 2004; Rienties et al., 2012; Torgerson & Torgerson, 2008). So in the next implementation of this module, using principles of Design-Based Research we could adjust the content of the message to encourage more learners to act upon the initial email, or change the amount and intensity of follow-up reminders. For example, we could adjust the narrative of the message (e.g., focussing more on social element of message), change the way we address the respective learner (e.g., Dear John, rather than Dear student), or provide some quotes from previous learners who struggled and found back-on-track website useful for their study to create a sense of relatedness (Bienkowski et al., 2012; Siroker & Koomen, 2013). By tracking learners' behaviour over the following two weeks, we can afterwards determine whether the quasi-experimental intervention was more or less successful in altering the behaviour of learners in comparison to the quasi-experimental control condition (i.e., the initial implementation). Finally, we could compare the academic performance difference between the new cohort and previous cohort in order to determine the impact on retention. However, a natural limitation of this kind of research (like in Scenario 1) is that the composition of learners in the follow-up study might be different in terms of attitudes, behaviour and cognition, and the environment in which they study may have changed (e.g., different VLE tools, teachers, support, funding structure).

Scenario 3 A switching replications design

An alternative scenario might be to conduct a specific alteration of the quasi-experimental study, namely a switching replications design, whereby first half of the cohort (Group A) will get the newly phrased intervention email in week 3, and the other half of the cohort (Group B) will receive the same email but only in week 4. In this way, Group A forms the intervention group during week 3, and can be compared and contrasted with the control Group B in terms of their behaviour, as illustrated in Figure 3. For example, if 33 learners in the intervention group access the study skills modules in week 3, and only 14 learners in the control group access the skills modules, we can argue that the impact of the intervention is that 19 more learners followed-up with the expected behaviour (arrow 1 in Figure 3). In week 4, 45 learners in Group A accessed the skills module, and after receiving the same intervention 35 learners in Group B accessed the website (arrow 2 in Figure 3). After five weeks, 35 learners in Group A accessed the skills module, and 44 in Group B (arrow 3 in Figure 3). As all learners received the support and more or less the same number of learners engaged with the website until the end of the module, we have not disadvantaged Group B in giving them the delayed feedback. However, we are able to state that the intervention will lead to an increase in access to the skills website of approximately 30 learners in a two week time period (both for Group A and Group B).

Figure 3 Quasi-experimental intervention with time-delay



Scenario 4 A/B testing within one study

A fourth scenario could be to implement A/B testing (Siroker & Koomen, 2013), whereby both groups get a similar treatment at the same point in time but the content/look-and-feel/navigation of the message is altered. For example, group A gets the intervention message in week 4 primarily phrased on cognitive dimension (e.g., did you know that students who accessed the back-on-track website were 23% more likely to pass the module?), while group B gets an altered intervention message containing a personal example of a previous learner (e.g., did you know that 23 students looked at the back-on-track website last week, and found the website extremely useful. For example, Mary from Liverpool said that "I was a bit unsure whether I was putting enough time into the course as I have a busy working life and taking care of two my two lovely, but demanding kids at the same time. The back-on-track website gave me feedback that I was well on track and gave me confidence that I am able to master this course)". Ideally both A/B interventions should be considered as educationally valuable/progressive and not to adversely disadvantage the educational experience of the "other" group.

If in comparison to the A-group 30 more learners in week 4 clicked on the follow-up link in the B-type message, we can conclude adding a personal example could activate (some groups of) learner behaviour. A/B testing is particularly useful to unpack and understand which types of interventions are appropriate for specific groups of learners. For example, perhaps mothers with children might be more inclined to follow the link in the B-type setting due to the narrative of Mary from Liverpool, while perhaps women without children or mean might actually be less inclined to engage as they cannot really relate to the story of Mary. Again we remind readers that when planning a particular intervention researchers need to be clear about which kinds of ABC effects they are trying to impact, how they are going to measure these effects, and which kinds of statistical approaches are going to be used to verify/reject these hypotheses.

Scenario 5 Randomised Control Trial

A final scenario could be to a full randomised control trial (RCT, Rienties et al., 2014; Slavin, 2008; Torgerson & Torgerson, 2008), whereby for example we at random give 1/3 of the cohort an intervention mail with follow-up phone call one week after the mail is send, 1/3 of the cohort only the intervention mail, and 1/3 of the cohort a placebo (e.g., email with non-task related message: "University cycling team is raising money for Cancer UK research, could you help?") or no specific intervention. By tracking the behaviours of learners in the two experimental conditions in comparison to the learners in the control condition, we should be able to determine the causal relations of the type and intensity of the intervention. In particular by linking these three interventions with ABC we would be able to determine for which groups of learners the additional phone call might have a positive effect (e.g., learners with low self-efficacy, anxiety, lack of engagement with VLE) and for whom it might lead to unexpected negative effects or "mothering" (e.g., highly active learners in VLE, intrinsically motivated learners). However, a natural limitation of RCTs is that substantial time and effort needs to be invested in order to plan, design, implement and evaluate these kinds of studies, which may not always be possible when a quick intervention is needed. This approach may be particularly useful in cases where there is no or mixed prior evidence that an intervention promises, or should be expected to be beneficial (and so inclusion of a placebo can be educationally justified).

Case-study 2 Helping learners-at-risk identified by predictive modelling

Our second case-study is an example of highly sophisticated learning analytics system developed by the OU, which uses a range of advanced statistical and machine learning approaches to identify learners potentially at-risk. In an introductory science module, data about 1730 learners were monitored using OU Analyse. The objective of the OU Analyse is to predict learners-at-risk (i.e., lack of engagement, potential to withdraw) in a course presentation as early as possible so that cost effective interventions could be made. For this module, the accuracy of predictions grew from about 50% at the beginning of the presentation to more than 90% at the end of the module presentation. Recall was stable at around 50%, but dropped to about 30% at the very end due to the incomplete results of preceding assessments. In OU Analyse, predictions are calculated in two steps:

- predictive models are constructed by machine learning methods from legacy data recorded in the previous presentation of the same course and,
- performance of learners is predicted from the predictive models and the learner data of the current presentation (Wolff et al., 2014; Wolff et al., 2013).

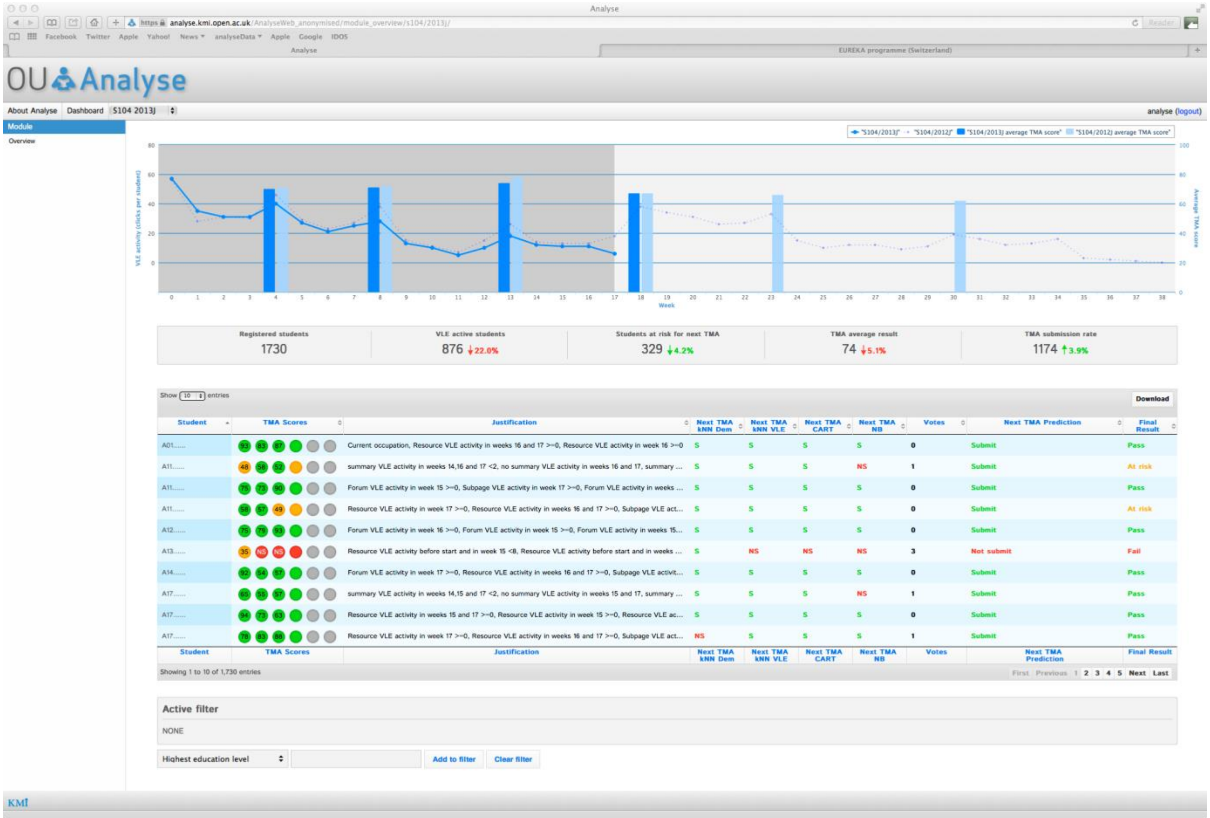
Machine learning methods aim at constructing predictive models that capture from legacy data patterns typical for succeeding, failing or withdrawing in formative/summative assessments and in the course. Two types of data were used for predictive modelling: demographic/static data and learner interactions with the VLE system. Demographic/static data include age, previous education, gender, geographic region, Index of Multiple Deprivation score, motivation, how many credits the learner is registered for, number of previous attempts on the course etc. VLE data represent a learner’s interaction with on-line study material and VLE interactions are classified into *activity types* and *actions*. Each activity type corresponds to an interaction with a specific kind of study material. For example, *resource* activity type typically refers to retrieving a segment of course text, OU *content* is used to point to the assessment, etc. (Wolff et al., 2014; Wolff et al., 2013).

These data were collected daily, but the OU Analyse algorithms used weekly aggregates. OU Analyse applies information theoretic criteria to select 4-6 activity types most informative for the outcome of the next assessments and for the final result. These activity types were used to build predictive models. Moreover, the frequency of learners' use of activities with selected activity types indicated which study material learners visited and how many times. Activity types that were not used pointed to a potential gap in knowledge and were used by the models as an input for individualised study recommender.

OU Analyse employs three machine learning methods to develop four predictive models: Bayesian classifier; Classification and regression tree; k Nearest Neighbours with demographic/static data; and k-NN with VLE data. These four models take into account different properties of data and complement each other. Each model independently classifies each learner into classes: will /will-not submit next assessment and will fail or pass the course. The final verdict of the prediction is done by combining the outcomes and using voting techniques of all four models (Wolff et al., 2014).

A list of learners likely not to submit the next assessment is sent every week to the module team. Results of learners' prior assessments (already known at the time of prediction) and demographic/static data were included. In Figure 4, a module view shows the average performance of the whole cohort and lists results of all learners with a traffic light symbols and brief justification of conclusions. Since the trajectory of each learner's activity types through presentation up to the current point of time is recorded, it can be used to recommend the best study material to successfully complete the assessment.

Figure 4 Predicting modelling of learners at-risk in OU Analyse in week 17



As indicated in Figure 4, OU Analyse predicted that 329 learners were “at-risk” before the fourth summative assessment point. The first part of graph in Figure 4 highlights average user engagement with the VLE, and compares this to previous implementations of the same module (dashed line). This indicated that engagement in week 17 was substantially lower than the previous implementation. OU Analyse also indicates the average assessment score on the second Y-axis, whereby average scores for the fourth assessment were predicted to be lower than those of the third assessment (and in comparison to the previous implementation). The lower part of Figure 4 gives a traffic light overview of each individual learner, and whether (or not) a learner is considered at-risk. For example, the first learner passed the three previous summative assessments with high grades and was predicted to do well on the forthcoming assessment as well. The second learner barely passed the first assessment, did slightly better on the second and third assessment, but was still characterised at-risk to pass the module as this learner had not engaged activity with the various VLE activities in week 14, 16 and 17. The sixth (and final) learner listed in Figure 4 had failed the first assessment, did not submit assessment 2-3 and was predicted not to submit assessment 4 and not to pass the module.

In order to allow researchers to evaluate the impact of interventions, we propose four different scenarios.

Scenario 1 Quasi-experimental follow-up

Based upon the experiences of 2014 and principles of Design-Based Research, the module chair of the module could redesign some of the learning activities after the third assessment, because a substantial group of learners in the previous implementation seemed to become less engaged at this point (as highlighted by lower VLE activity and lower assessment scores for fourth assessment). For example, qualitative learner evaluation feedback and input from tutors may indicate that one of the two textbooks used for this time period (week 13-18) was considered to be difficult and too abstract. As a result, the module chair could, for example, change this textbook with a more accessible, interactive online textbook with ample practices and real-world examples how principles of physics could be applied. In OU Analyse, we would be able to compare VLE activity of learners in week 13-18 with the previous implementation. More importantly, OU Analyse would be able to track each individual learner and determine whether their predictions of success will change (or not) due to this intervention.

Scenario 2 A switching replications design

An alternative scenario might be to conduct a switching replications design study, whereby for example half the cohort would start in week 13-15 with the original textbook and its respective tasks (Group A), while the other half of the cohort (Group B) would start with the new textbook. In week 16 the groups swap, whereby Group A would continue with the new textbook, while Group B would continue with the original textbook. In this way, both groups get the same two textbooks and related tasks, but in a different order so that the impact of the different textbooks on attitudes, behaviour and cognition can be compared and contrasted.

Scenario 3 A/B testing within one study

A third scenario could be to implement A/B testing (Siroker & Koomen, 2013), whereby both groups get a similar treatment but Group A, for example, starts in week 14 with an interactive exercise using an embedded video-quiz in the interactive textbook, followed by a theoretical part, and concluded with a short formative test, while Group B starts with the same quiz but in a text-based format. This would allow us to track whether providing embedded video-quizzes leads to more engagement with the theoretical part and cognition.

Scenario 4 Randomised Control Trial within one study

A final scenario could be to a full randomised control trial, whereby one third of the cohort at randomly gets the new textbook with interactive assignments, one third of the cohort receives the new textbook with text-based assignments, and finally one cohort gets the original textbook. In this way, we can test whether the new textbook leads to a more engaged learning behaviour and cognition, and whether the level of interactivity encourages or hampers rich learning. Given that OU Analyse incorporates a range of attitudinal and demographic data, this would also allow us to determine the impact of these three conditions for specific groups of learners.

Discussion and Learning Analytics Intervention and Evaluation Framework

Substantial progress in learning analytics research has been made in recent years to predict which groups of learners are at-risk (Agudo-Peregrina et al., 2014; Calvert, 2014; Gasevic et al., 2013; Macfadyen & Dawson, 2010). However, we argue that the largest challenge for learning analytics research and practice still lies ahead of us: using learning analytics modelling, which types of interventions have a positive impact on learners' Attitudes, Behaviour and Cognition (ABC). Two embedded case-studies in social science and science were discussed to illustrate some notions of how evidence-based research approaches could be used in learning analytics, namely comparison with previous implementations, quasi-experimental research, a within-quasi-experimental research, A/B-testing, and randomised control trials.

Each of these five scenarios has unique affordances and limitations. For academics familiar with educational research and who have sufficient data interpretation skills, in particular the first three scenarios are relatively straightforward to implement using principles of Design-Based Research (Collins et al., 2004; Rienties & Townsend, 2012). For academics who do not have these skills, educational psychologists, learning & teaching specialists or data-interpreters might help them to make informed suggestions for follow-up interventions (Clow, 2014; Rienties et al., 2012). By collecting as much data as possible from a range of sources, and by triangulating quantitative and qualitative results, academics and teachers can use data from previous and current implementations to identify bottlenecks in the learning design and how this influences ABC of learners. Afterwards, a design-based intervention (Collins et al., 2004) would allow academics and teachers to test, verify, compare and contrast whether (or not) the expected changes in ABC of learners indeed materialised.

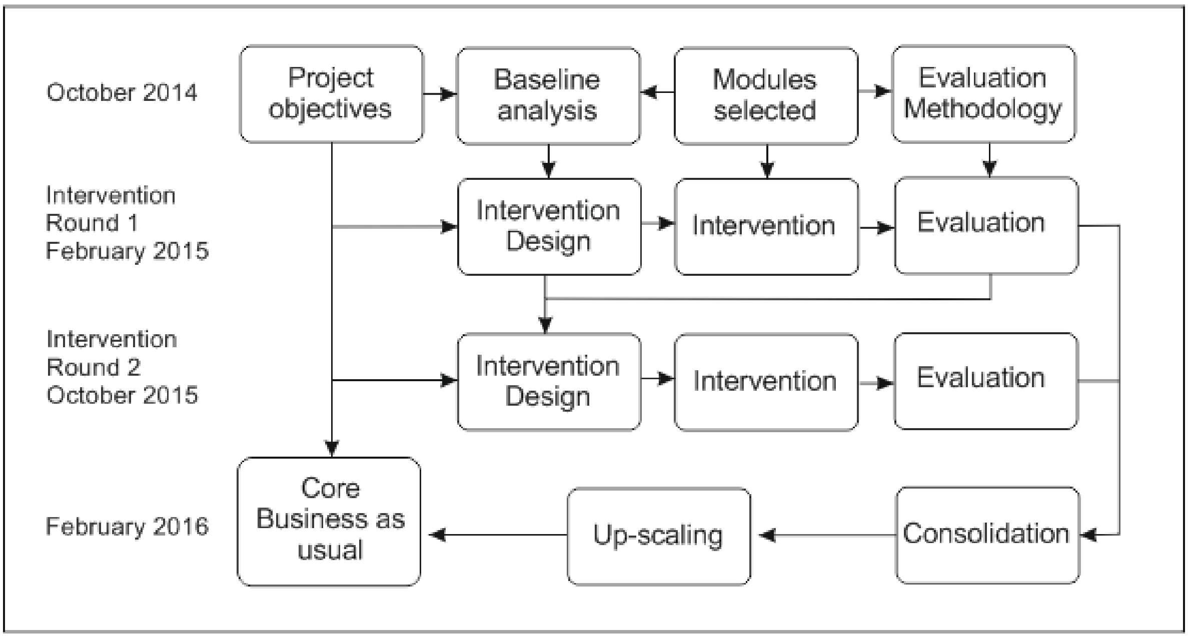
Implementing A/B testing or RCT-testing, the "gold-standard" in research (Slavin, 2002, 2008; Torgerson & Torgerson, 2008), is partly more complex due to organisational, technical, and (potential) ethical barriers. For example, not all VLE systems allow teachers to randomly assign learners to two or three different groups, and use subsequent adaptive routing to track whether learners in the experimental condition(s) behave differently than those in the control condition. Even if IT systems allow for A/B testing or adaptive routing, substantial manual labour may be needed to assign learners to the different conditions. In particular with relatively small samples (<200), even random assignment in different conditions might not guarantee an equal distribution of learner characteristics (ABC) across the conditions, so researchers may need to check appropriate sampling. Finally, obtaining ethical permission to conduct A/B testing or RCT may not always be straightforward, and at times unpractical or unethical (depending on the proposed intervention). Nonetheless, in line with Slavin (2008) we argue that often only with RCTs and A/B testing can we provide robust and reliable evidence under which conditions learning analytics can provide cost-effective, yet rich interventions to our students.

Implementing a Learning Analytics Intervention and Evaluation Framework

The Open University is currently implementing a Learning Analytics Intervention and Evaluation Framework (LA-IEF) with 15 large cohort first year modules across the various disciplines. If organisations like the OU are going to adopt and continually finance learning analytics approaches, we need to provide evidence-based results where we can identify learners at-risk (e.g., using OU Analyse), deliver (personalised) intervention suggestions that work, and most importantly interventions that are cost-effective. Therefore, one pragmatic reason for choosing first year modules at the OU is that retention rates amongst these learners are traditionally lower than in follow-up years (Calvert, 2014; Richardson, 2012a). By using the power of learning analytics where it is most needed, but across a range of disciplines, we expect to be able to provide an evidence-based approach under which conditions particular interventions are successful in altering learners' ABC.

As illustrated in Figure 5, using principles of Design-Based Research (Collins et al., 2004; Rienties & Townsend, 2012) extensive dialogue with key stakeholders (e.g., module chairs, tutors, librarians, multi-media designers, IT, learners) are being conducted between in September 2014 – April 2015 as a baseline study to determine what is going well and what bottlenecks are (potentially) present in each of these 15 modules according to these stakeholders. At the same time, these modules will be extensively evaluated using a range of learning analytics approaches (e.g., OU Analyse, VLE monitoring) and existing evaluation practices with the OU, thereby leading to a solid base-line study.

Figure 5 LA-IEF framework as implemented at the OU



Follow-up discussions with module chairs and relevant stakeholders in November-December 2014 using the insights of learning analytics will determine which types of interventions will be implemented in the next implementation of the modules in February 2015 in an evidence-based manner. For some modules, a quasi-experimental design will be used, whereby based on the results of the baseline study (parts of) the module will be altered. For other modules, we aim to use A/B testing or RCT testing to be able to directly identify cause and effect of particular interventions. More importantly, by planning, implementing and evaluating these interventions across a range of disciplines, the LA-IEF model will help to advance methodological robustness of learning analytics research, by comparing and contrasting research findings across different domains and contexts using an evidence-based approach.

A crucial element of the Learning Analytics Intervention and Evaluation Framework is the recognition that most interventions and innovations lead to unexpected, possibly even negative results. While we expect that several of the 15 interventions will lead to positive impacts on ABC of learners, several interventions will have no (measurable) impact on attitudes, behaviour or cognition of (groups of) learner, or perhaps even lead to (in)direct negative effects. 40 years of educational research has highlighted that learning and cognition is inherently complex (Arbaugh, 2005, 2014; Hattie, 2009; Richardson, 2012a; Rienties et al., 2012; Slavin, 2008), but only with a clear evidence-based research programme will researchers be able to unpack and understand under which conditions we can help learners-at-risk in particular contexts. Given that in education there can conceivably be several different ways of teaching that may potentially be equally effective, then the question of A-B testing/randomised trials does not have to be between haves and have-nots or small variation; it could be between learning design A and learning design B where there is some grounds to expect each to be effective (a multiplicity or range of potential designs that carry a similar risk of failure). Indeed, flipping the entire question from 'how often does it work' to 'how often does it fail' may lead us to see learning design as an exercise in risk minimisation.

A continuous cycle of interventions in the next two years are planned at the OU, which will help to replicate, fine-tune and generalise for those interventions that had an initial positive effect. For those interventions which did not lead to a positive effect, taking the metaphor from medical science further fine-tuning the doze, level, and type of interventions will lead to a robust understanding what works and what does not. By a continuous cycle of planning, designing, implementing, and evaluating interventions across a range of modules, module chairs and the wider organisation will be empowered to embed these intervention "recipes" into their practice. Finally, by moving towards an evidence-based research approach to learning and teaching, we aim to move towards a robust, flexible and cost-effective university-wide implementation of learning analytics which will reduce drop-out and allow learners to reach their full potential.

Acknowledgements

We would like to thank Prof Belinda Tynan, Kevin Mayles, and Avinash Boroowa from the Learning and Teaching Centre at the Open University UK for their continuous support, and critical feedback to the LA-IEF framework.

References

Agudo-Peregrina, Á. F., Iglesias-Pradas, S., Conde-González, M. Á., & Hernández-García, Á. (2014). Can we predict success from log data in VLEs? Classification of interactions for learning analytics and their relation with performance in VLE-supported F2F and online learning. *Computers in Human Behavior, 31*(February), 542-550. doi: 10.1016/j.chb.2013.05.031

Aguiar, E., Chawla, N. V., Brockman, J., Ambrose, G. A., & Goodrich, V. (2014). *Engagement vs performance: using electronic portfolios to predict first semester engineering student retention*. Paper presented at the Proceedings of the Fourth International Conference on Learning Analytics And Knowledge, Indianapolis, Indiana.

Arbaugh, J. B. (2005). Is there an optimal design for on-line MBA courses? *Academy of Management Learning & Education, 4*(2), 135-149. doi: 10.5465/AMLE.2005.17268561

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

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

Clow, D. (2014). *Data wranglers: human interpreters to help close the feedback loop*. Paper presented at the Proceedins of the Fourth International Conference on Learning Analytics And Knowledge.

Collins, A., Joseph, D., & Bielaczyc, K. (2004). Design Research: Theoretical and Methodological Issues. *The journal of the learning sciences, 13*(1), 15-42.

Conde, M. Á., & Hernández-García, Á. (2015). Learning analytics for educational decision making. *Computers in Human Behavior, 47*(0), 1-3. doi: 10.1016/j.chb.2014.12.034

Ferguson, R. (2012). Learning analytics: drivers, developments and challenges. *International Journal of Technology Enhanced Learning, 4*(5), 304-317. doi: 10.1504/ijtel.2012.051816

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.

García-Peñalvo, F. J., Conde, M. Á., Alier, M., & Casany, M. J. (2011). Opening learning management systems to personal learning environments. *Journal of Universal Computer Science, 17*(9), 1222-1240. doi: 10.3217/jucs-017-09-1222

Gasevic, D., Zouaq, A., & Janzen, R. (2013). "Choose your classmates, your GPA is at stake!": The association of cross-class social ties and academic performance. *American Behavioral Scientist, 57*(10), 1460-1479. doi: 10.1177/0002764213479362

Giesbers, B., Rienties, B., Tempelaar, D. T., & Gijselaers, W. H. (2013). Investigating the Relations between Motivation, Tool Use, Participation, and Performance in an E-learning Course Using Web-videoconferencing. *Computers in Human Behavior, 29*(1), 285-292. doi: 10.1016/j.chb.2012.09.005

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

Greller, W., & Drachsler, H. (2012). Translating Learning into Numbers: A Generic Framework for Learning Analytics. *Journal of Educational Technology & Society, 15*(3).

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.

Inkelaar, T., & Simpson, O. (2015). Challenging the 'distance education deficit' through 'motivational emails'. *Open Learning: The Journal of Open, Distance and e-Learning*, 1-12. doi: 10.1080/02680513.2015.1055718

Jindal-Snape, D., & Rienties, B. (Eds.). (2016). *Multi-dimensional transitions of international students to higher education*. London: Routledge.

Knight, S., Buckingham Shum, S., & Littleton, K. (2013). *Epistemology, pedagogy, assessment and learning analytics.* Paper presented at the Proceedings of the Third International Conference on Learning Analytics and Knowledge.

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

Martin, A. J. (2007). Examining a multidimensional model of student motivation and engagement using a construct validation approach. *British Journal of Educational Psychology, 77*(2), 413-440. doi: 10.1348/000709906X118036

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

Nistor, N., Baltes, B., Dascălu, M., Mihăilă, D., Smeaton, G., & Trăuşan-Matu, Ş. (2014). Participation in virtual academic communities of practice under the influence of technology acceptance and community factors. A learning analytics application. *Computers in Human Behavior, 34*, 339-344. doi: 10.1016/j.chb.2013.10.051

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.

Papamitsiou, Z., Terzis, V., & Economides, A. (2014). *Temporal learning analytics for computer based testing*. Paper presented at the Proceedings of the Fourth International Conference on Learning Analytics And Knowledge, Indianapolis, Indiana.

Pekrun, R., Goetz, T., Frenzel, A. C., Barchfeld, P., & Perry, R. P. (2011). Measuring emotions in students' learning and performance: The Achievement Emotions Questionnaire (AEQ). *Contemporary educational psychology, 36*(1), 36-48. doi: 10.1016/j.cedpsych.2010.10.002

Pintrich, P. R., & De Groot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology, 82*(1), 33-40.

Richardson, J. T. E. (2012a). The attainment of White and ethnic minority students in distance education. *Assessment & Evaluation in Higher Education, 37*(4), 393-408. doi: 10.1080/02602938.2010.534767

Richardson, J. T. E. (2012b). The role of response biases in the relationship between students' perceptions of their courses and their approaches to studying in higher education. *British Educational Research Journal*, *38*(3), 399-418. doi: 10.1080/01411926.2010.548857

Rienties, B., & Alden Rivers, B. (2014). Measuring and Understanding Learner Emotions: Evidence and Prospects *LACE review papers* (Vol. 1). Milton Keynes: LACE.

Rienties, B., Giesbers, B., Tempelaar, D. T., Lygo-Baker, S., Segers, M., & Gijselaers, W. H. (2012). The role of scaffolding and motivation in CSCL. *Computers & Education, 59*(3), 893-906. doi: 10.1016/j.compedu.2012.04.010

Rienties, B., Giesbers, S., Lygo-Baker, S., Ma, S., & Rees, R. (2014). Why some teachers easily learn to use a new Virtual Learning Environment: a Technology Acceptance perspective. *Interactive Learning Environments*. doi: 10.1080/10494820.2014.881394

Rienties, B., Toetenel, L., & Bryan, A. (2015). *"Scaling up" learning design: impact of learning design activities on LMS behavior and performance.* Paper presented at the 5th Learning Analytics Knowledge conference, New York.

Rienties, B., & Townsend, D. (2012). Integrating ICT in business education: using TPACK to reflect on two course redesigns. In P. Van den Bossche, W. H. Gijselaers & R. G. Milter (Eds.), *Learning at the Crossroads of Theory and Practice* (Vol. 4, pp. 141-156). Dordrecht: Springer.

Siemens, G. (2013). Learning Analytics: The Emergence of a Discipline. *American Behavioral Scientist*, *57*(10), 1380-1400. doi: 10.1177/0002764213498851

Siemens, G., Dawson, S., & Lynch, G. (2013). *Improving the quality of productivity of the higher education sector: Policy and strategy for systems-level deployment of learning analytics*: Solarresearch.

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

Stiles, R. J. (2012). Understanding and Managing the Risks of Analytics in Higher Education: A Guide (pp. 1-46): Educause.

Tempelaar, D. T., Niculescu, A., Rienties, B., Giesbers, B., & Gijselaers, W. H. (2012). How achievement emotions impact students' decisions for online learning, and what precedes those emotions. *Internet and Higher Education, 15*(3), 161–169. doi: 10.1016/j.iheduc.2011.10.003

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.

Wise, A. F. (2014). *Designing pedagogical interventions to support student use of learning analytics*. Paper presented at the Proceedings of the Fourth International Conference on Learning Analytics And Knowledge, Indianapolis, Indiana.

Wolff, A., Zdrahal, Z., Herrmannova, D., Kuzilek, J., & Hlosta, M. (2014). *Developing predictive models for early detection of at-risk students on distance learning modules, Workshop: Machine Learning and Learning Analytics* Paper presented at the Learning Analytics and Knowledge (2014), Indianapolis.

Wolff, A., Zdrahal, Z., Nikolov, A., & Pantucek, M. (2013). *Improving retention: predicting at-risk students by analysing clicking behaviour in a virtual learning environment.* Paper presented at the Proceedings of the Third International Conference on Learning Analytics and Knowledge.

Yin, R. K. (2009). *Case study research: Design and methods* (Vol. 5): Sage.