Open Research Online

The Open University's repository of research publications
and other research outputs

Analytics for Tracking Student Engagement

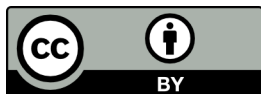## Journal Item

For guidance on citations see FAQs.

Version: Version of Record

Link(s) to article on publisher's website:
http://dx.doi.org/doi:10.5334/jime.590

oro.open.ac.uk

ARTICLE

Analytics for Tracking Student Engagement

Christine Gardner, Allan Jones and Helen Jefferis

Although there has been much research in the area of data analytics in recent years (e.g. Shum & Ferguson 2012), there are questions regarding which analytic methodologies can be most effective in informing higher education teaching and learning practices (Gibson & de Freitas 2016).

This project focuses on one module within the School of Computing and Communications in the STEM faculty of The Open University, UK, to gain a clearer understanding on why students might, or might not, engage with computer aided learning and teaching (CALT) resources. We explore the use of specific CALT resources on the module 'Communications Technology', a print-based module with a range of online resources designed to supplement the text. In particular we explore the possible correlation between use of CALT resources and student examination performance.

The research questions cover two key areas; the effectiveness of the analytics tools and students' perception of the CALT resources. Data analytics were used to determine when students engaged with the CALT resources and whether this was at predicted times during the module. Student feedback via interview was used to explore what motivates students to engage with CALT resources, whether students understand a topic more deeply as a result of using CALT resources, and if students are deterred if the resources are too complicated or time consuming.

Our conclusion from this case study is that learning analytics are useful for tracking student engagement. The analytics were very useful to review during module presentation, specifically for analysing students' online behaviour. The supplementary interviews helped to shed light on the potential significance of the data gleaned.

1. Introduction

The UK's Open University (OU) has evolved significantly since its creation fifty years ago and has developed its own style of distance learning, 'supported open learning', offering students opportunities to study flexibly, whether at home, work, library or other study centre. Before the advent of the Internet, students relied solely on printed study materials. Key to its continuing success is the utilisation of new technologies.

This study looks at the use of learning analytics to uncover student engagement with computer aided learning and teaching (CALT) resources in the Open University module TM355 Communications Technology. This module is an elective component in the university's honours degree in Computing and IT. The module covers such topics as radio propagation, digital signal modulation, source coding, error control, optical fibres, DSL broadband and mobile communications. Parts of the module are supported by sophisticated CALT resources, particularly in relation to coding and error control.

The module is studied towards the end of the students' degree level studies and introduces several complex topics. To aid study of such material, additional experiential learning (Kolb 1984) is available via online interactive activities, designed to supplement the written materials. These are referred to within the printed materials and are added to the students' study planner, grouped together to make them relatively easy to find. An example is shown in **Figure 1** below.

The research was motivated by a particular examination question used in the 2017 examination. The question, on the topic of error control, was in a part of the examination paper where students had a choice of questions to answer. The question related to techniques of error control that had been taught in print, and demonstrated interactively with a CALT resource which students were strongly advised to use, but could not be compelled to use. In this study learning-analytics data was used retrospectively to investigate the use of the relevant CALT resource by students who chose to answer the question, and as an aid to framing interview questions relating to the use of CALT resources. Rather than using the analytics data to predict student performance, this retrospective use of the data appears to us to be a novel use of learning analytics data, to further investigate what has happened.

Open University, GB

Corresponding author: Christine Gardner (c.f.gardner@open.ac.uk)
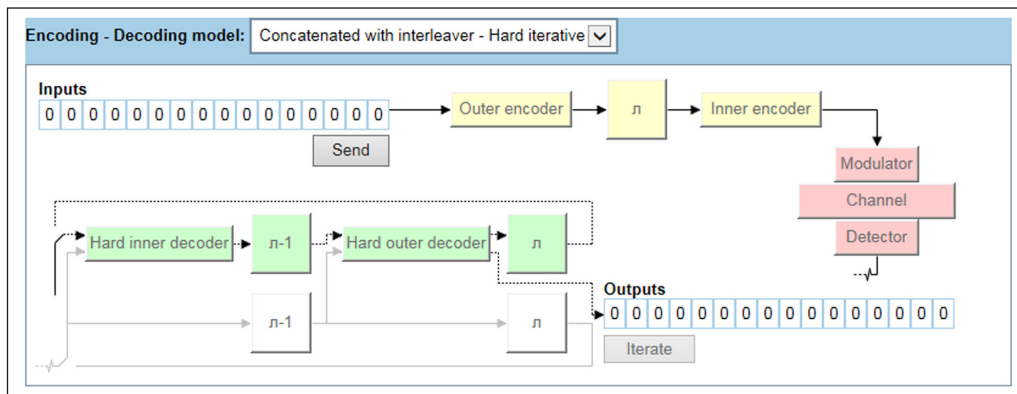
Figure 1: Example of TM355 CALT resource (online activity).

The hypothesis that the research team aimed to test is that those students who engaged with the CALT resources tended to perform better at exam time. Much time and resource had been invested in producing the CALT resources to supplement the printed materials, so there is a question on how extensively they are being used and, if not being used, why that might be.

In particular, the research covered two key areas; the effectiveness of the analytics resources and students' perception of the CALT resources.

Via data analytics we reviewed:

- when the students engaged with the CALT resources and whether this was at predicted times during the module;
- whether students revisited the CALT resources.

Via individual student feedback we explored:

- what motivates students to engage with CALT resources;
- whether students understand a topic more deeply as a result of using CALT resources;
- if students are deterred if the resources are too complicated or time consuming.

2. Learning Analytics

Learning analytics, in George Siemens's widely used definition, are the 'measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environment in which it occurs' (quoted in Bodily & Verbert 2017: 405). The gathering and use of statistical data about learners is not new. For example, pass/fail rates and grade distributions have long been a tool of educators and educational researchers, but 'learning analytics' connotes something more than traditional performance statistics.

The development of learning analytics has largely been an outgrowth of the development of virtual learning environments (VLEs), where students' online study is conducted in a computer-based learning environment. Such an environment allows the students' progress, performance of tasks, use of resources, etc. to be recorded. Although such data could be used for monitoring an individual student, generally learning-analytics data is aggregated from many students (sometimes hundreds or thousands) in order to identify significant trends or patterns of study behaviour.

Although there has been much research in the area of data analytics in recent years (e.g. Shum & Ferguson 2012), there are questions regarding which analytic methodologies can be most effective in informing higher education teaching and learning practices (Gibson & de Freitas 2016).

Indeed, learning analytics have also been seen as an outgrowth of the development of the 'big data' movement. (Kop, Fournier, & Durand 2017; Littlejohn 2017). Two 'big data' applications of learning analytics have received particular attention. One is the use of learning analytics to predict students' behaviour or success. (for example, Slater & Baker 2019). The other is the use of learning analytics in conjunction with learning design to help with module revision and improvement. (Sclater, Peasegood, & Mullen 2016). Slater and Baker (2019) point out a potential problem with the predictive approach, which is its tendency to assume that learning is continuous and incremental, allowing extrapolations to be made. This ignores the possibility of learning being discontinuous, in which the gaining of a sudden insight produces an unpredicted shift in performance. An additional problem with the 'predictive' approach is that although it can reveal correlations, causal connections between student activity and educational progress remain unclear. The problematic nature of predictive learning analytics possibly underlies a trend identified by Viberg et al. (2018: 108), who report that research in learning analytics in higher education is shifting from prediction towards 'a deeper understanding of students' learning experiences.' The authors of the present paper see their work as part of this trend with analytics data being reviewed retrospectively to help gain further insight into student performance.

Two issues in particular are common to both the 'predictive' approach and the 'student experience' approach, both of which are equally relevant to our retrospective approach. The first of these concerns the ethics of monitoring students. Siemens (2019) refers to concerns around student privacy in connection with learning analytics and Bodily and Verbert (2017) observe that some types of learning analytics potentially reduce student autonomy as teachers and administrators increasingly become framed as managers of learners.

The second issue concerns what analytics actually represent. Analytics data typically comprises counts of mouse-clicks made by students (and their date and time) to access particular virtual learning environment (VLE) pages or web pages, numbers of messages posted (along with date and time), and, possibly, time spent on a particular VLE page or web page. Records of clicks and time spent therefore serve as proxies for learning activities and resource use, and possibly as rather poor proxies.

For instance, Macfadyen and Dawson (2010) found that time spent on educational resources, as indicated in learning analytics data, did not correlate with academic performance. However, their study did find a strong correlation between the number of messages a student posted and their eventual grade. Even where a strong correlation can be demonstrated, though, causation cannot be assumed. Do students who disproportionately engage with a particular aspect of a VLE do so because they are good students, or do they become better students by doing so? Thus students' motivation for clicking on particular resources, the degree of attention students pay to them, and the results of doing so cannot reliably be deduced from analytics data. To investigate issues of motivation and attention supplementary techniques such as surveys and interviews must be used. In the present case, learning analytics data was supplemented with interviews. This appears to the authors to be a novel, or at any rate uncommon, use of learning analytics.

3. Method

The trigger for implementing this research was to test the hypothesis that students could gain higer marks when examined on key topics if they had engaged fully with the related CALT resources. In order to rigourously test this it was necessary to trace the behaviour of individual students. The analytics tool used presents information at cohort level; and although the underlying information relating to individuals is stored in the data gathered from the VLE, it is not displayed. We therefore had to use special procedures to recover information about individual students.

The research questions cover two key areas; the effectiveness of the analytics tool and students' perception of the CALT resources. The methodology employed a mix of quantitative and qualitative research methods, in particular the collection of data analytics and use of semi-formal interviews.

Via data analytics it was possible to review when the students engaged with the CALT resources and whether at predicted times during the module. It was also possible to collect data to establish whether students revisited the CALT resources. Via individual student telephone interviews a more in-depth view could be established regarding what motivates students to engage with CALT resources, whether students understand a topic more deeply as a result of using CALT resources, or if students are deterred if resources are too complicated or too time consuming.

3.1. Data collection – analytics for action (A4A)
The main data analytics tool selected for the research was Analytics for Action, A4A (Hidalgo 2018). A4A can provide detail of how students are engaging with specific online materials. Data is presented at a high level, with the aim of providing a module-level analysis of how students are engaging with online materials. The framework for application of A4A has six phases, as shown in **Figure 2**, that can help module teams continually review and improve student experience by identifying actions to be taken.

A4A is a visual platform, providing a summary of student performance using real-time data. For example, **Figure 3** depicts student interaction with a specific online resource that relates directly to assessed material. The vertical axis represents the number of students per week engaging with a CALT resource, the horizontal axis represents the
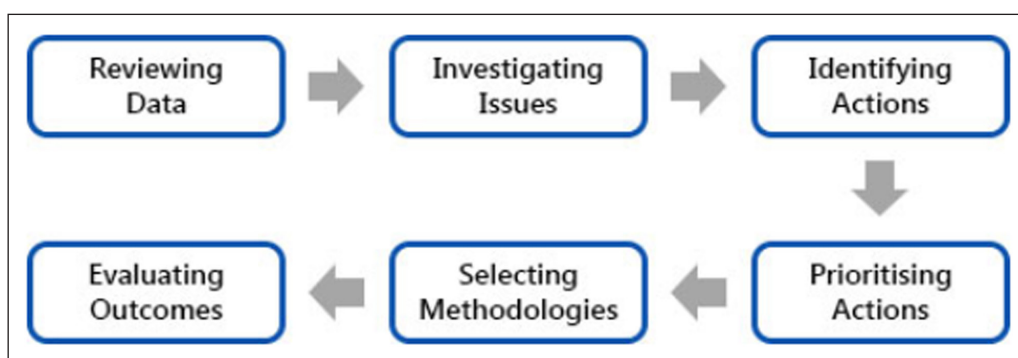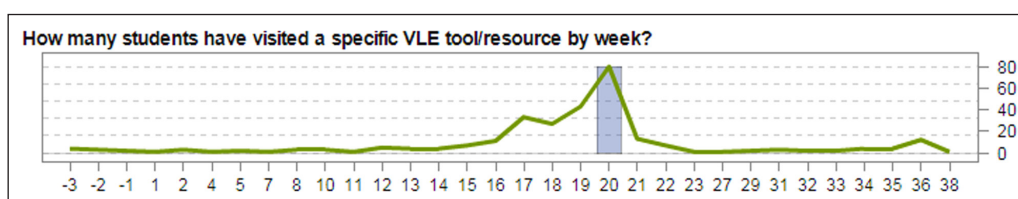

Figure 2: A4A six phases.

Figure 3: Hamming codes – Online resource predominantly used in relation to assessment.

study weeks of the module and the grey vertical bar represents an assignment due date. It can be seen that peak use of the related online resource ties in with the submission of the second assignment, due in week 20.

In **Figure 4**, the use of the CALT resource has a different kind of pattern, as it is used predominantly during Block 2 of the module, between the first and second assignments (represented by the first and second vertical bars). There is also a small peak at the end of the module, suggesting that some students return to the online resource at revision time. However, the number of students engaging per week is relatively low, considering the cohort size of over 300 students.

The A4A data can help a module team make evidence-based decisions, with the ultimate goal of improving student experience on that module (Evans, Hidalgo, & Calder 2017). A limit to the usefulness of the A4A dashboard is that in its usual format it cannot identify online activity at an individual student level. In consultation with the analytics design team it was established that the underlying data could be presented at an individual student level if required, but this required bespoke procedures that could not be made routinely available. The required data was made available, and the use made of it is outlined in the following section.

3.2. Phases of research and ethical considerations
For this study, a sub-set of 48 students from the cohort was selected. These were all the students who answered a particular 2017 exam question on error correction. The question was not answered well, resulting in a poor average mark for this question. Individual student activity on the related CALT resource for these 48 students was collected and mapped alongside their exam performance.

In order to investigate possible reasons for the poor performance on this question, research was conducted, in three main phases. During the first phase (2017 to July 2018) a pilot study was conducted which commenced

with the collection of key analytics data via A4A on CALT resource use during the relevant module presentation. From this the sub-set of 48 students was identified for further research. In consultation with the analytics team this data was interrogated more deeply to establish the patterns of CALT use of these 48 students. The second phase was designed to supplement the analytics data via semi-formal interview questions, to help address limitations of analytics data such as those noted earlier. All students from the cohort were invited to interview. Interviews took place in July 2018 and were followed by an initial review of findings. The third phase involved action and dissemination of findings. The research was approved by the Open University's Research Ethics Committee and has been logged as GDPR compliant.

## 4. Results and Discussion on Recommendations
The following is a summary of findings from both the data analytics and student interviews.

4.1. Results from A4A regarding error coding activity relating to exam
In reviewing the data relating to the TM355 students who answered the specific exam question relating to error coding, the small sample of 48 students was selected from the cohort of 329 students who sat the final examination. The data relating to their online activity with the related CALT resource was mapped alongside their examination score for the question (**Table 1**).

This snapshot relating to student performance suggests that those who engaged fully with the CALT resource did relatively well. Naturally correlation does not necessarily mean causation (Ferguson & Clow 2017).

From **Figure 4** it can be calculated that between study weeks 9 and 20 inclusive about 212 students used the CALT resource relating to error control codes. This was calculated by adding the number of visits per week, between weeks 9 and 20. Some of these might be students who
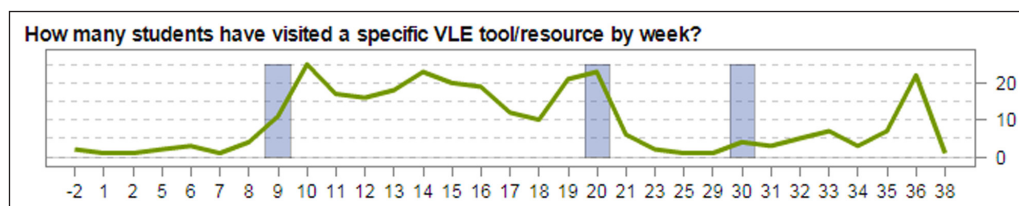
Figure 4: Error control codes – Online resource used at specific times during a module.

Table 1: Student examination performance comparisons.

Student category (within the 48-student subset)	Average examination question score
All students who answered the question	45% – bare pass (pass mark 40%)
Did not use error control codes CALT resource	30% – fail
Used error control codes CALT resource during the module at least once	53% – pass
Used error control codes CALT resource specifically at revision time	52% – pass
Used error control codes CALT on more than one specific date i.e. returned to package	58% – clear pass

used the resource more than once, so they are double counted in that figure of 212. Even so, although the use of the resource might be seen as disappointing, it would be reasonable to suppose that within the cohort of over 300 students at least half the students used it in this period. That is also consistent with the 53% usage figure for students who attempted the examination question relating to error control codes.

However, far fewer than half of the student cohort felt confident enough to do the related exam question. In the relevant part of the exam paper, the students select two out of three questions to answer. If students distributed themselves evenly across the three exam questions, we would expect 67% of them to attempt each question. In fact, the 48 students who attempted the question related to error correction represented 15% of the cohort.

It is reasonable to suppose that by the time the students sat the examination they had not used the online resource for several months, unless they also used it in the revision period. From the A4A data it can be seen that the weekly use after week 20 was very low. Between weeks 20 and 38 there were about 61 uses of the CALT resource, with only about 22 in the week before the exam. It is therefore reasonable to hypothesise that students did not appreciate the importance of including the CALT resources in their revision.

To check whether the 48 students who answered the error-correction question (and on average performed poorly) were relatively weak students relative to the cohort, their general performance was compared with that of the 271 students from the cohort who chose not to answer the question. This was done by comparing their cumulative achievement on previous modules with that of the cohort. This is summarised in **Figure 5**.

In **Figure 5**, the data set was coded as original_sample=1 for those who answered the error-correction question, and the others flagged original_sample=0, so differentiating between the groups.

A comparison of the mean probability of passing this module in each group shows they are very similar in predicted performance. The sample group of 48 students is marginally weaker at (approximately) 0.85 pass probability compared to (approximately) 0.87, as highlighted above, but not to the extent that they should achieve significantly lower results. This suggests that even though the sample size is small there is no underlying reason to assume that the sample group should perform any differently to the rest of the cohort and it does not explain their poor performance on this question. In other words, students who attempted this question were representative of the whole cohort. Their poor performance on the error-correction question indicates there was something anomalous about this question. The most obvious anomaly was that it was taught both via text and CALT resource. In the opinion of the question's author, although the question did not require the students to have used the CALT resource in order to be able to answer it, the CALT resource gave useful insights into the topic, alongside practice in answering this type of question. For the research team, therefore, a question that presents itself is why some students do not use the CALT tools as they study the material.

4.2. Interview results

Two sets of interviews were conducted, one for the pilot study (3 students) and one for the following cohort (5 students). Students were selected via the university's student research project panel, inviting them to contribute to the research. An email request was sent to all students eligible to take part, and all of those who responded were interviewed. Results were based on both pilot study and main cohort responses.

The Open University engages in distance learning so face-to-face interviews were not possible, as the students are widely distributed across the UK and beyond. For the pilot study the intial plan was for interviews to be conducted via the Open University's Skype for Business system. However, during the interview period a restriction on recording external calls became evident, so interviews and recordings were completed via mobile phone. For the second cohort, interviews were conducted via Adobe Connect using its inbuilt recording facility. This had the added bonus of a visual screen, on which the online descriptions were added as an aide-memoir for the interviewees. The interviews were semi-structured.

Several key points were highlighted by students during the interview process. For example, some students noted that the CALT resources were very good for self-testing. It was also noted that the activities provided a different way to learn rather than text and they were visual and interactive. As an example, the benefit of being able to step forwards and backwards through an animation in order to go back and check things was commented upon.

Summary Statistics

Results

The MEANS Procedure

Analysis Variable : P_target_result1 Probability of student passing the module							
original_sample	N Obs	Mean	Std Dev	Minimum	Maximum	Mode	N
0	271	0.8670889	0.1349558	0.2279104	0.9876422	0.5742115	271
1	48	0.8516501	0.1430410	0.1702492	0.9850042	0.5742115	48

Figure 5: Statistical analysis of sample.

"Seeing the coding in practice and having an inter-action helped".

A particular benefit relating to complex themes was noted, as the related CALT resources supplemented the written text, thus helping the students to understand the topic more fully by interacting with the online version of the materials.

"The sequence of coding and decoding is explained in the book and this is done quite well, but the activities help you to do that for real and allow you to apply the theory".

It was also noted that the online resources were useful in providing a high-level summary, as large sections of the printed materials (2 or 3 pages) could be summarised in a paragraph or two of the online activity. This reinforces the suggestion of highlighting use of the CALT resources for revision, where students need to best utilise the time available. An issue noted was the lack of information regarding the estimated time for engaging with the online activities. The notional time needed varied between activities and this could not be determined unless actually engaging with the activity.

Although student perception of the CALT resources was generally positive there were also more negative responses, for example uncertainty regarding the estimated timings for the activities, alongside some lack of activity clarity, for example the following quotation relating to the 'launching a wave' activity.

"It was hard to understand the direction of the dipole and how it was radiating waves".

The student found it difficult to work out what was happening. A slightly different animation showing which direction was which could have helped the student's understanding, so this is a further idea to progress.

4.3. Discussion on recommendations and actions

As mentioned in connection with **Figure 2**, the A4A framework was designed to help educators improve their online tuition, and was therefore adopted in order identify possible improvements in the module. The initial review of A4A data revealed that some students were using the online activities effectively to support their study of printed module materials, although many students did not fully engage. A particular issue was highlighted after the 2017 examination, so further analysis of the data was undertaken which suggested that those students who utilsed the CALT resource associated with the examina-

tion question performed slightly better than those who did not, although these results should be treated with some caution. A series of student interviews was conducted to gain further insight into their perception of the use of these activities.

The findings from this study suggest that there are several actions that could be taken, for example:

· give a clearer indication of time needed for the CALT activities (although obviously this will vary for each student);
· add short descriptions about what kind of activity it is, for example interactive, video;
· promote the activities in a new module introductory or revision video or podcast;
· use the module forums to promote the activities further;
· have 'talking heads' of students saying how useful they were;
· add further detail to the introduction to certain activities, for example to explain the orientation in the 'launching a wave' activity;
· tell students not to overlook the CALT activities when revising;
· engagement with the online materials should be encouraged, as they should enhance the learning experience by exposing students to a wider variety of learning techniques.

Several of these ideas suggested via the interviews have already been implemented and others could be actioned in the future. For example, **Figure 6** depicts a section of a resource that has been produced to give students an overview of the activity type and typical timings, alongside a direct link to the activity and an indication on where it fits in the student study calendar.

Also, a new revision podcast has been produced which specifically promotes the use of the CALT resource at revision time, and will hopefully result in more students revisiting the online resources.

5. Conclusions

Data analytics can prove useful in analysing student performance and in modifying a module in the light of what is revealed. However, as with any statistical data, interpretation is required. Data analytics do not 'speak for themselves'. As mentioned earlier in connection with the literature on analytics, the analytics data yielded by the online learning environment is not, by itself, illuminating. Establishing the significance of analytics data is likely to require the use of additional strategies, of which interviews are an example. The example discussed here revealed some

Block	Part	Week	Online Activity	Activity Name and module link	Short Description	Estimated Time Req.
1	1	1-3	1.1	Fourier Transforms	Interactive activity showing time-domain representation and frequency-domain representation for sine, square, sawtooth and triangular waveforms.	15-30 mins

Figure 6: TM355 CALT resource description and timing example.

of the limitations of aggregated data. Knowing that a certain percentage of a student cohort did not do a particular activity could raise an alarm about the activity, but typically one would need to know more about the group of students identified. In what ways might they be representative or unrepresentative of the cohort? Pursuing this question is likely to require drilling down to data about individual students, which can be (as in the present case study) beyond what the analytics tool is intended to do, and might raise ethical concerns We see here another version of the 'prediction versus student experience' dilemma faced by designers and users of data analytics tools.

In the present case study, follow-up interviews revealed the puzzling inconsistency that the CALT resources are considered useful but are under-utilised, particularly during the revision period. This fact indicates a clear course of action for the creators of the module, which is to urge students not to confine their revision solely to the printed texts, which contain the bulk of the teaching material. A revision advice podcast, newly introduced, stresses this and gives other revision advice.

Our general conclusion from this case study is that learning analytics have undoubtedly proved useful for tracking student engagement, but have required a certain amount of 'hand-crafting' to extract additional information that is not routinely available, and supplementary interviews to shed light on the potential significance of the data gleaned. In particular, we have found analytics useful as a retrospective tool, rather than a prospective tool, for analysing what happened during a presentation, and whether the resources were used as intended or to their fullest extent.

Competing Interests
The authors have no competing interests to declare.

References

Bodily, R and **Verbert, K.** 2017. Review of research on student-facing learning analytics dashboards and educational recommender systems. *IEEE Transactions on Learning Technologies*, 10(4): 405–418. DOI: https://doi.org/10.1109/TLT.2017.2740172

Evans, G, Hidalgo, R and **Calder, K.** 2017. Analytics for action – enabling in-presentation evidence-based change on OU modules. *Quality Enhancement Report Series,* Issue No 2017/7.

Ferguson, R and **Clow, D.** 2017. Where is the evidence? A call to action for learning analytics. *LAK'17 Proceedings of the 7th International Learning Analytics and Knowledge Conference* (pp. 56–65). Vancouver, BC. DOI: https://doi.org/10.1145/3027385.3027396

Gibson, D and **de Freitas, S.** 2016. Exploratory analysis in learning analytics. *Technology, Knowledge and Learning*, 21(1): 5–19. DOI: https://doi.org/10.1007/s10758-015-9249-5

Hidalgo, R. 2018. Analytics for action: Using data analytics to support students in improving their learning outcomes. In: Ubachs, G and Konings, L (eds.), *The envisioning report for empowering universities* (2nd ed). Maastricht, Netherlands: European Association of Distance Teaching Universities. pp. 6–8.

Kolb, DA. 1984. *Experiential learning: Experience as the source of learning and development.* Englewood Cliffs, NJ: Prentice Hall.

Kop, R, Fournier, H and **Durand, G.** 2017. A critical perspective on learning analytics and educational data mining. In: Lang, C, Siemens, G, Wise, A and Gašević, D (eds.), *The Handbook of Learning Analytics. Society for Learning Analytics Research.* pp. 319–326. DOI: https://doi.org/10.18608/hla17.027

Littlejohn, A. 2017. Learning and work: Professional learning analytics. In: Lang, C, Siemens, G, Wise, A and Gašević, D (eds.), *Handbook of Learning Analytics. Society for Learning Analytics Research.* pp. 269–277. DOI: https://doi.org/10.18608/hla17.023

Macfadyen, LP and **Dawson, S.** 2010. Mining LMS data to develop an "early warning system" for educators: A proof of concept. *Computers and Education*, 54: 588–599. DOI: https://doi.org/10.1016/j.compedu.2009.09.008

Sclater, N, Peasegood, A and **Mullen, J.** 2016. Learning analytics in higher education. *JISC Report.* Available at https://www.jisc.ac.uk/reports/learning-analytics-in-higher-education.

Shum, SB and **Ferguson, R.** 2012. Social learning analytics. *Educational Technology and Society*, 15: 3–26.

Siemens, G. 2019. Learning analytics and open, flexible and distance learning. *Distance Education*, 40(3): 414–418. DOI: https://doi.org/10.1080/01587919.2019.1656153

Slater, S and **Baker, R.** 2019. Forecasting student master. *Distance Education*, 40(3): 380–394. DOI: https://doi.org/10.1080/01587919.2019.1632169

Viberg, O, Hatakka, M, Bälter, O and **Mavroudi, A.** 2018. The current landscape analytics in higher education. *Computers in Human Behavior*, 2018: 98–110. DOI: https://doi.org/10.1016/j.chb.2018.07.027