# Where is the evidence?
# A call to action for learning analytics

Rebecca Ferguson
Institute of Educational Technology
The Open University
Walton Hall, Milton Keynes,
MK7 6AA – UK
Rebecca.Ferguson@open.ac.uk

Doug Clow
Institute of Educational Technology
The Open University
Walton Hall, Milton Keynes,
MK7 6AA – UK
Doug.Clow@open.ac.uk

## ABSTRACT
Where is the evidence for learning analytics? In particular, where is the evidence that it improves learning in practice? Can we rely on it? Currently, there are vigorous debates about the quality of research evidence in medicine and psychology, with particular issues around statistical good practice, the 'file drawer effect', and ways in which incentives for stakeholders in the research process reward the quantity of research produced rather than the quality. In this paper, we present the Learning Analytics Community Exchange (LACE) project's Evidence Hub, an effort to relate research evidence in learning analytics to four propositions about learning analytics: whether they support learning, support teaching, are deployed widely, and are used ethically. Surprisingly little evidence in this strong, specific sense was found, and very little was negative (7%, N=123), suggesting that learning analytics is not immune from the pressures in other areas. We explore the evidence in one particular area in detail (whether learning analytics improve teaching and learners support in the university sector), and set out some of the weaknesses of the evidence available. We conclude that there is considerable scope for improving the evidence base for learning analytics, and set out some suggestions of ways for various stakeholders to achieve this.

## Categories and Subject Descriptors
K.3.1 [**Computers and Education**]: Computer Uses in Education

K.4.1 [**Computers and Society]**

## Keywords
Access, Ethics, Evidence, Evidence Hub, Generalisability, Learning Analytics Cycle, Reliability, Validity

## 1. INTRODUCTION
The first Learning Analytics and Knowledge conference in 2011 explored what the call for papers described as 'the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs'. In contrast to other areas of technology-enhanced learning (TEL) and quantitative educational research, there was a concern with 'closing the loop' [19] to achieve improvements in learning practice. The learning analytics community and literature have grown steadily since then. How far have we progressed towards that goal and how can we evidence this progress?

To answer these questions, we will first explore how evidence has developed in practice in two entirely separate fields (medicine and psychology). This overview shows that there are problems with evidence in many scientific fields and that many of the problems we encounter (for example, publication bias, the Hawthorn Effect and confusion between causality and correlation) are not confined to learning analytics. We then move on to an examination of the use of evidence in education before focusing on the case of learning analytics and suggesting possible actions.



## 2. WHAT IS EVIDENCE?

### 2.1 Evidence-based medicine
In the 1970s, Cochrane raised concerns about the evidence base for medical practice. This led to the establishment of the Cochrane Collaboration (now simply 'Cochrane')[1] in 1993, with the aim of improving the evidence base for practice. This dovetailed with the development of the evidence-based medicine movement, 'the conscientious, explicit and judicious use of current best evidence in making decisions about the care of individual patients' [58].

**Table 1: Example of a hierarchy of evidence [51]**

| |
|---|
| Systematic reviews and meta-analyses |
| Randomised controlled trials with definitive results |
| Randomised controlled trials with non-definitive results |
| Cohort studies |
| Case-control studies |
| Cross sectional surveys |
| Case reports |

Although the evidence-based medicine movement stresses that 'evidence-based medicine is not restricted to randomized trials and meta-analyses' [58], these are its main focus, because these methods rank high in hierarchies of evidence. Like the hierarchy in Table 1 above, they rank evidence in terms of reliability, with case reports considered the least reliable and randomised controlled trials (RCTs) the most reliable. This is recognised to be

[1] http://www.cochrane.org/

a simplistic view that can prove problematic in contexts such as social or public health [51], but the intention is to focus on the strongest possible evidence to support the evidence-based medicine that is now almost ubiquitous in Western healthcare.

Although concerns were raised about this approach, there was also considerable optimism. More recently, there have been worries that the idea has been diverted from its original goals [33]. In addition, more fundamental concerns about the quality of the underpinning research evidence are coming to light.

The ethics of insisting on RCTs of treatments 'known' to be effective are complex and still under debate. On the one hand, there are many examples of treatments that were 'known' to be effective that turned out to be actively harmful; but on the other, insisting on the highest quality of evidence before taking action can cause significant avoidable harm. There are at least now established procedures for ending trials early when sufficiently strong evidence has been gathered.

One issue is the use of 'surrogate endpoints' [53]. An example is blood pressure: we know that high blood pressure is a risk factor for cardiovascular mortality, so it might seem reasonable to assess a new drug on the basis of whether it lowers blood pressure, particularly as mortality rates are generally low, so a large trial would be needed to evaluate the effect on mortality. However, it may be that the drug lowers blood pressure but does not affect mortality, or has severe adverse effects that outweigh any benefit.

Another major issue is publication bias, whereby uninteresting or negative findings are not reported. Positive results are more likely to be written up and accepted as publications, while negative results are more likely to languish, unloved, in file drawers, creating a 'file drawer effect'. This is a particular concern in the area of clinical research, where it has been the practice of some pharmaceutical companies not to publish all the research they have conducted in to the safety and efficacy of new treatments. There is a movement to address this, with the ambitious AllTrials[2] project working to have 'All trials registered, all results reported'.

Ioannidis made the bold claim that most published research findings are false [37], substantiating this with a model of the research process that is not restricted to medicine. He later argued that most of the true research that is published is not useful in clinical practice [38]. There is a strong statistical rationale underpinning this concern, along with a complex set of incentives for researchers and publishers, which are perhaps most vividly illustrated by recent issues in the broad field of psychology.

## 2.2 Evidence in psychology

Concerns were raised in psychology about an over-reliance on samples of people from western, educated, industrialized, rich, and democratic (WEIRD) societies, and how representative those samples were of humanity as a whole [35].

This concern about external validity and a deeper concern about internal validity sparked a 'replication crisis'. After attempts to replicate famous psychological results failed, a high-profile Reproducibility Project repeated 100 key correlational studies to see if the same results could be obtained. The results were disappointing. 'A large portion of replications produced weaker evidence for the original findings', with only 36–47% of replications succeeding, depending on the measure chosen [48]. These efforts have been highly controversial, with critiques of studies often posted in the grey literature (chiefly blog posts) and on social media.

An underlying issue is the use of statistics, including 'researcher degrees of freedom' to make a study reach significance [61]. This is an issue if researchers carry out multiple comparisons but only report the significant ones. It also arises where 'researchers can perform a reasonable analysis given their assumptions and their data, but had the data turned out differently, they could have done other analyses that were just as reasonable' [29]. An underlying problem is that any research carried out with low pre-study odds is prone to false positives [37].

The problems are deep-seated. A '60-year meta-analysis of statistical power in the behavioural sciences [shows] that [statistical] power has not improved despite repeated demonstrations of the necessity of increasing power' [62]. Incentives prompt researchers and journals to produce findings that are interesting and these drive high false discovery rates, even if replications are commonplace [62].

## 2.3 Evidence in education

In the case of education, it is extremely challenging to carry out RCTs at all. The contexts in which learning occurs are highly variable and personal, and there is even less consensus about the ethics of conducting trials than there is in medicine.

Prominent efforts around this form of evidence include the What Works Clearinghouse[3] in the USA, the work of the Gates Foundation in K12 education[4], and a position paper on 'Building evidence into education' published by the UK Government [31] written by a prominent advocate of evidence-based medicine.

There is broad consensus even among advocates of RCTs in education that they are not a panacea [30] and 'cannot sit in a research vacuum' [2]; others are considerably less supportive of these approaches.

The transfer of research evidence into practice is also far from perfect. This problem is perhaps most vividly illustrated by the example of learning styles. Despite comprehensive evidence against the concept, all but a tiny minority of practising teachers, across cultures, believe that 'individuals learn better when they receive information in their preferred learning style' [36].

One can legitimately criticise many studies in education, and in technology-enhanced learning (TEL), for insufficient rigour. However, a narrow focus on methodological purity and RCTs is unlikely to prove productive. A more important issue is the framing of the research question and a rigorous consideration of the goals and underlying model of learning and teaching [42]. Good-quality quantitative research needs to be supported by good-quality qualitative research: we cannot understand the data unless we understand the context.

The issue of surrogate end points also arises in education. Even if results can be validly compared, there is often a lack of consensus that the test measures what people want education to achieve. Nonetheless, standardised testing has been carried out extensively.

## 2.4 International data gathering

Countries worldwide collect evidence about education that they can use to inform and assess social and educational policies, as well as to judge their performance in relation to other countries.

[2] http://www.alltrials.net/

[3] http://ies.ed.gov/ncee/Wwc/

[4] http://k12education.gatesfoundation.org/

The methods they use to do this are usually rooted in psychometrics, the science of psychological assessment [57]. Well-known examples of systematic evidence gathering include the Programme for International Student Assessment (PISA) and the Trends in International Maths and Science Survey (TIMMS).

PISA is an international survey that aims to evaluate education systems worldwide by testing the skills and knowledge of 15-year-old students. Since 2000, every three years, students from randomly selected schools worldwide have taken tests in reading, mathematics and science. These two-hour tests combine open-ended and multiple-choice questions that are based on a real-life situation. Information collected through questionnaires filled in by schools and students provides context for the test results.

TIMMS is designed to enable participating countries to make evidence-based decisions for improving educational policy. Since 1995, the survey has been used to monitor trends in mathematics and science achievement every four years. Its assessments provide data relating to student performance in different domains of mathematics and science as well as problem solving in each of these areas. The studies also collect data about some of the contextual factors that affect learning, including school resources, student attitudes, teaching practices and support at home.

The results of these studies are widely cited and are used to influence the education policy of many countries. They are also open to criticism on the grounds of validity, reliability and generalisability. The US-based National Education Policy Center [13] reviewed the main critiques that have been made of such international tests. It noted that students in the samples from different countries are not directly comparable. When the figures are adjusted to take this into account, countries that have made large gains on TIMMS appear to have made little or no gains on PISA, which calls into question the validity of the results. The error terms of the test scores are large, so the accuracy of figures can be questioned. Despite the large sample sizes, the students tested are not representative samples of a country's population, so the results are not necessarily generalizable.

Validity does not depend only on how well tests are designed and validated, but also on how defensibly the resulting evidence is employed [17]. As the results of these tests are influential, it is tempting for countries to undermine the validity of their results by gaming the system. 'Although there are formal design recommendations and sampling protocols, participation of schools and classrooms tends to be locally determined based on priorities, politics and resources – especially in developing nations' [17].

The results of these tests may also be interpreted in ways that are not valid. Commentators and politicians confuse correlation and causation when they claim direct causal connections between test scores and aspects of schooling. This misuse of evidence is also seen when measurements of attainment in numeracy or literacy are taken to provide evidence of the effectiveness of teachers, schools, states or even the country as a whole [43].

Each of these tests is centrally controlled so that a consistent research design is employed across countries and across time. The tests and their results are open to public scrutiny, and the results of this scrutiny can be fed back into the tests in order to increase the value of the evidence that they provide.

## 2.5 Evidence in learning analytics

In the case of an entire research field, such as learning analytics, it is much more difficult to move consistently towards evidence that is generalizable, valid and reliable. A major problem for learning analytics is that the field is now so diverse it is impossible for any individual or team to keep up with all the literature.

The development of literature reviews helps but these tend to be aimed at researchers and not practitioners. The LAK Dataset[5] makes machine-readable versions of literature available. This is a rich resource, but is not easily accessible by readers. The SoLAR website brings resources together in its Info Hub.[6] These provide a useful introduction, but only for those with time to explore a wide range of resources.

A different way of dealing with the problem of making evidence accessible was developed in the field of open education. In 2011, the Open Learning Network (OLNet) project launched the Evidence Hub for Open Education. The aim was to provide an environment that could be used to represent the collective knowledge of the Open Education community. The Evidence Hub could be used to investigate the people, projects, organisations, key challenges, issues, solutions, claims and evidence that scaffold the Open Education movement [25].

# 3. DEVELOPING A LEARNING ANALYTICS EVIDENCE HUB

The Learning Analytics Community Exchange (LACE) project[7] has used the model developed by the OER Research Hub to produce an Evidence Hub for the learning analytics community.

## 3.1 Developing Evidence Hub criteria

In the case of an Evidence Hub, the term 'evidence' refers to the available body of facts or information that indicates whether a particular proposition is true or valid. In order for learning analytics resources to be classified as evidence, they therefore need to relate to a proposition that may be true or false.

Work to identify the propositions that would underpin the LACE Evidence Hub began with the structured development of a framework of quality indicators [59]. An initial version of this framework included five criteria, each associated with four quality indicators [59]. LACE consortium members refined initial propositions based on these. This resulted in four propositions:[8]

**A:** Learning analytics improve learning outcomes.

**B:** Learning analytics improve learning support and teaching, including retention, completion and progression.

**C:** Learning analytics are taken up and used widely, including deployment at scale.

**D:** Learning analytics are used in an ethical way.

As some evidence is valuable but does not have a positive or negative polarity in relation to these propositions, the Evidence Hub allows evidence to be classified as positive, negative or neutral in relation to a proposition.

These propositions were introduced to the wider learning analytics community at an event that attracted over 400 participants from across Europe. Groups discussed the propositions and existing evidence for or against them. These discussions showed that the

[5] https://solaresearch.org/initiatives/dataset/

[6] https://solaresearch.org/core/

[7] http://www.laceproject.eu/

[8] http://www.laceproject.eu/evidence-hub

propositions could be used effectively to structure evidence in the field of learning analytics.

# 4. LACE EVIDENCE HUB

The LACE team has worked to ensure that the Evidence Hub includes as much relevant evidence as possible. Different work streams within the project – schools, universities and workplace learning – each contributed evidence related to its sector. A focused literature search examined papers from early LAK conferences. The Hub was publicized at learning analytics events and, in 2016, those submitting a paper to LAK were invited to link it to the Evidence Hub criteria. This link to LAK meant that coverage of 2016 is more extensive than in previous years. The Evidence Hub does not provide complete coverage – it is currently skewed towards papers written in English, for LAK, since 2015. Importantly, its focus does not include most of the literature around intelligent tutoring systems. Nevertheless, it represents the most comprehensive and systematic coverage of evidence that is currently available – and its open nature means that anyone can contribute additional evidence.

## 4.1 LACE Evidence Hub findings

One of the early findings was the surprising quantity of published research papers in the LAK Dataset that did not contain evidence in this strong sense of being evidence for or against one of the four broad propositions. Many of the LAK papers are not empirical research. Of those that are, some are evidence only of intermediate effects (e.g. reliability of predictions of at-risk students) rather than evidence for one of the propositions (e.g. that this can improve their learning, Proposition A). The Evidence Hub took a fairly broad view of whether a piece of research met this criterion, as explored in section 4.2 below.

At the time of writing,[9] the LACE Evidence Hub contains 123 pieces of evidence, summarised in Table 2.

**Table 2: Summary of positive (+), negative (-) and neutral (±) contents of Evidence Hub in relation to four propositions**

| | Supports Learning | Supports teaching | Deployed at scale | Deployed ethically | Total |
|---|---|---|---|---|---|
| Schools | +10 -2 ±3 | +7 -1 ±5 | +5 -0 ±2 | +1 -2 ±0 | 38 |
| Universities | +14 -0 ±5 | +22 -0 ±6 | +6 -2 ±1 | +3 -1 ±0 | 60 |
| Workplace | +8 -0 ±2 | +2 -0 ±0 | +0 -0 ±0 | +0 -0 ±0 | 12 |
| Informal | +3 -0 ±1 | +0 -0 ±0 | +2 -0 ±0 | +0 -0 ±0 | 6 |
| Cross Sector | +0 -0 ±1 | +0 -0 ±1 | +3 -0 ±0 | +0 -1 ±1 | 7 |
| Total | 47 | 44 | 21 | 9 | 123 |

It is immediately clear that, although this is the seventh annual LAK conference, there is still very little hard evidence about learning analytics. What is more, the evidence that we do have is significantly skewed towards the positive: only 7% of the findings are negative. This issue that was examined in detail at the 2016 LAK Failathon [21]. As discussed in section 2.1, publication bias is a well-known problem in medicine and in most other empirical disciplines. The many accounts of failures given at the Failathon suggest that the mainly positive evidence presented in the learning analytics literature does not fully represent the findings of research work within the discipline.

The papers in the Evidence Hub mainly relate to learning and teaching in schools and universities. There are particular gaps in the evidence about informal learning, workplace learning, and ethical practice.

[9] October 2016

## 4.2 Evidence problems in one sector

We considered the situation in more detail by focusing on one area. The Higher Education sector of the Evidence Hub contains more evidence than any other. Most of that evidence relates to the proposition, '**Learning analytics improve learning support and teaching, including retention, completion and progression**'. This analysis therefore focuses on the 28 papers that have been classified as evidence that learning analytics improve teaching in universities. The majority of the evidence (22 items) is positive. Six items are neutral [6; 16; 22; 40; 46; 56] and there is so far no evidence against the proposition.

While this appears to be good news, it seems unlikely. With hundreds of researchers working across the world, surely one has tested a learning analytics innovation and found that it does not improve learning support and teaching [21]?

The Evidence Hub mapping tool suggests that work on supporting teaching in universities with learning analytics is being developed and widely disseminated by only a few institutions. Figure 1 shows that almost all the evidence related to this proposition (represented by green circles) originates in a handful universities in the south of Australia, the west of Europe and the north-east of the USA. (The green circle in the Pacific Ocean represents studies associated with more than one area.) Outside these main areas, there are single pieces of evidence (represented by orange pins) in Singapore and Greece. As the LAK conference was receiving submissions from 31 countries by 2013 [63] and has since grown considerably, we might have expected to see evidence being produced and disseminated in many more countries.

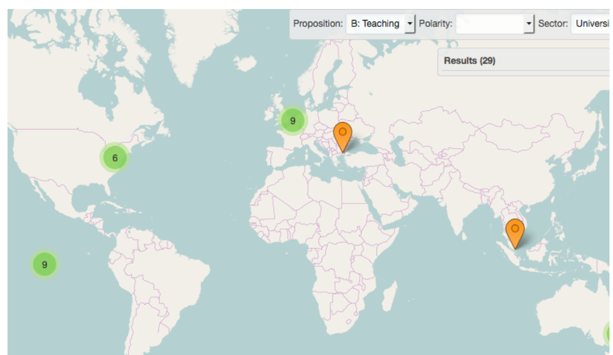


**Figure 1: Evidence related to universities and teaching**

The evidence that is available in the Evidence Hub falls into four main groups: evidence that can support institutions, evidence that can support the development of learner models, evidence with the potential to support teaching and evidence that has had impact on teaching and its effects. In the analysis below, sample sizes are noted for each study. Unless otherwise stated, the samples were made up of students from a single institution.

**Evidence that can support institutions**

Some evidence focuses on student support across the university. Two pieces of evidence fit into this category. The first relates to the Learning Analytics Readiness Instrument (LARI), designed to help institutions gauge their readiness to implement learning analytics. It is ready for use, but has not yet been deployed (N=560 respondents, 24 institutions) [49]. The second analyses the financial benefits for an institution of using an early alert system. Over three years, the system had significant financial benefits to the institution (N=16,124). Although the focus is on financial benefit to the institution, the implication is that significant numbers of students benefited by being supported to remain at the institution. However, no information is supplied about the early alert system or how it was deployed [34].

**Evidence that can support development of learner models**

Learner models represent information about a student's characteristics or state, such as their current knowledge, motivation, metacognition and attitude. Such models can be used to provide automated support for students. Three pieces of evidence deal with factors that could be incorporated within a learner model: time spent on task (N=259) [50], misconceptions about one-digit multiplication (N not specified) [64] and affect (N=44) [1]. A fourth piece of work on automated support focuses on detection and analysis of reflective writing (N=30 pieces of student work) [8]. While all these studies have the potential to improve learning support and teaching, there is no evidence as yet that they have actually done so.

**Evidence with potential to support teaching**

Some of the evidence that falls into this category has the potential to support teaching, but there is no clear route from research into practice. Topic modelling has potential as an analytic tool to help teachers assess reflective thoughts in written journals (N=80) [18]. An analytics dashboard designed for users of interactive e-books could potentially be used by teachers [41]. A rule-based indicator definition tool (RIDT) could support a personalized learning analytics experience (N=5 staff, 7 students) [47]. Studying eye fixation patterns could enable educators to understand how their instructional design using online learning environments can stimulate higher-order cognitive activities (N=60).

Other evidence has clearer pathways into practice and promises to have a positive effect on practice in the near future. A study of possible predictors of student success makes the important point that predictors are only useful in cases where intervention is possible (N=1,005 and N=1,006) [65]. Another study supports teachers build on learning analytics by introducing a conceptual framework designed to transform learning design into a teacher-led enquiry-based practice (N=12 teachers, 4 universities) [5].

Some data help predict student failure or drop-out. Changes in user activity in a virtual learning environment can predict failure when compared with their previous behaviour or that of students with similar learning behaviour (N=7,701) [66]. Analysis of data about student movement within and across a learning community can be used to develop strategic interventions in the learning of at-risk students. For example, teaching staff were found to be more commonly located in the networks of high-performing students and analytics made staff more aware of this (N=1,026) [23]. An investigation of individual student, organizational, and disciplinary factors that might predict a student's classification in an Early Warning System, as well as factors that predict improvement and decline in their academic performance, resulted in tentative recommendations for educators (N=566) [7].

Visualising data can make analytics more accessible to teachers. Visualising online student engagement/effort provides instructors with early opportunities for providing additional student learning assistance and intervention when and where it is required (N=1,026) [24]. Using visualisations produced by the Student Activity Meter tool can help awareness and understanding of student resource use and student time-spending behaviour (two case studies, N=12 and N=20 evaluators, mainly teachers) [32].

Two large-scale studies have produced robust findings and recommendations that are currently being put into practice. Learning design activities seem to have an impact on learning performance, in particular when modules rely on assimilative activities (N=19,322) [55]. Development of appropriate communication tasks that align with the learning objectives of the course appears likely to enhance academic retention (N=111,236) [54]. However, this work is too early in the implementation process to have produced clear evidence that learning support and teaching have improved. The implications of these studies remain tentative while a larger ongoing body of work investigates whether their recommendations work in practice.

The studies considered up to this point could, in future, have a positive impact on teaching and learner support. With the possible exception of the analysis of the financial impact of an early alert system [34], there is no clear evidence that they have already achieved that impact. Only two studies in the Evidence Hub provide evidence that analytics have prompted changes in teaching and support that have impacted on learners.

**Evidence of impact on teaching**

The first of these is the highly cited (242 citations in Google Scholar) work on the use of the Course Signals learning analytics system at Purdue University. The paper reported that courses that implemented Course Signals realized a strong increase in satisfactory grades, and a decrease in unsatisfactory grades and withdrawals. Students who participated in at least one Course Signals course were retained at rates significantly higher than their peers who did not (N=23,000 students, 140 instructors) [3]. This is an important study, with major implications for learning analytics as a field, and it is considered further in the next section.

The second study in this section used a predictive model for student drop-out that 'was very similar [… to] the predictive model developed at Purdue University' [44]. This study (N=1,379) reported a small but statistically significant difference in content mastery rates (C grade or above) between intervention groups and controls. It also found statistically significant differences in withdrawal rates between intervention groups and controls. Worryingly, students in the combined treatment group were more likely to withdraw than those in the control groups [44]. It seems possible – although this is not explored in the study – that the improvement in grades was due to weaker students dropping out. The authors note that the increased dropout rate is consistent with students 'withdrawing earlier in the course (as opposed to remaining enrolled and failing)'. They were not able to give data about failure rates, however, so their data are also consistent with the intervention encouraging weak students who would otherwise have completed successfully to drop out.

So, when considering the positive influence of learning analytics on teaching and learning support in the higher education sector – the area in which we appear to have the most evidence – our strongest example remains the 2012 work on Course Signals. It therefore makes sense to examine this evidence in detail and to ask whether it is valid and reliable.

## 4.3 Evidence problems in one study

Valid, reliable evidence is not easy to obtain in the field of TEL One reason is the 'Hawthorne Effect', popularly considered to be the change of behaviour by subjects of a study due to their awareness of being observed. In the original account of this effect, in 1925, workers in a study were found to increase productivity. This increase was not due to the variable under consideration but was, at least in part, because records were taken more frequently than usual, which amounted to increased supervision [39]. This effect may be exacerbated in classrooms, where studies typically attract extra resource that is removed once the trial is over.

In order to establish the reliability and validity of a study, it is important to understand both its basis and its context. The Course Signals tool was developed at Purdue University in Indianapolis. Work began with an exploratory study by Campbell [9], using course management system student data from 2005 to determine undergraduate success. Campbell's study, which was written up in his doctoral thesis and in Educause publications [9-11], suggested that these data could be used as an appropriate proxy for overall student effort. The university's use of data to identify at-risk students became an early example of what were then known as academic analytics [10; 12], and the university's early-warning system was developed into a tool called Course Signals [52].

Course Signals used empirical data to build a student-success algorithm. This considered past academic performance but placed more emphasis on student effort and help-seeking. When students were classified as at-risk, this classification triggered interventions set up by instructors [52]. By 2009, more than 7,000 students were using the system, and the results reported in 2012 appeared very promising. Courses that implemented Course Signals saw an increase in A and B grades and a decrease in lower grades. In addition, students who participated in at least one Course Signals course appeared to be retained at rates significantly higher than their peers, and students who took two or more courses with Course Signals were consistently retained at rates higher than those who had only one or no courses with Course Signals [3].

The evidence seemed strong. Research based on five years of data with thousands of students showed that analytics could help to improve grades and increase retention. These were important claims that inspired many researchers and institutions to engage with work on learning analytics. Others took a more detailed look at the results that had been reported, and their critiques began to appear online in blogs. In August 2012, Caulfield raised the point that, between 2007 and 2009, retention at the university had also risen substantially for courses that did not employ Course Signals. This suggested that university-wide changes were having a significant effect on retention figures [14]. Caulfield followed this with a blog post the following year in which he asked whether the study had controlled for the number of classes a student took, and how the first to second year retention had been calculated [15].

Caulfield's overarching concern was whether students had been retained because they had taken more courses that used Course Signals, or whether they took more of those courses because they had been retained. Essa built a simulation to explore this issue and blogged that correlation had indeed been confused with causation [27]. Clow suggested, again in a blog post, that 'the Purdue researchers should urgently re-analyse their data, taking world-class statistical advice […] and publish the outcome in raw, unreviewed form as fast as possible, and then write it up and submit it for peer review' [20].

So, on the one hand, the learning analytics community has a peer-reviewed conference paper that is frequently cited and that has inspired many. On the other hand, we have a serious challenge to the methodology the paper employed. This critique appears in the 'grey literature' (publications that are not peer reviewed) and is less commonly cited. The evidence provided by the peer-reviewed paper is therefore in doubt. It seems probable that Course Signals does have a positive effect on students, even if this is simply an effect on their grades on courses that run Course Signals but, without running another analysis, we cannot be sure.

As Clow suggests, the obvious course would be to reanalyze the data and to open that analysis to public scrutiny. However, the authors were university staff rather than faculty members and were not free to continue the study without university approval. They are now working on different projects or at different institutions and do not have the data access or the resources to carry out another analysis. They are therefore in the unenviable position of seeing their work called into question without being in a position to amend, extend or defend their analysis. We would like to make it quite clear that what we are presenting here is a critique of the Course Signals study as published. It is in no way a personal attack on the two authors, whom we know to be talented and dedicated learning analytics researchers. The statistical issue here is not trivial and the apparent error is entirely understandable.

One view is that it is not in the interests of Purdue University to re-examine its data because 'the university is effectively making money on the strength of research claims that have now been called into question' [28], and it continues to make those claims without re-examining them [45].

Another view is that it might not be in the interests of the learning analytics community to dig too far into the data that underpins one of its flagship examples. However, that is exactly what we do need to do, because we need to build our work on firm foundations.

## 4.4 Comparison to other areas

It should be emphasised again that this issue is not one unique to learning analytics. The pattern here – an exciting, significant finding in a published paper, which appears out to have major issues that are discussed only in the grey literature – is a very common one in the 'replication crisis' in psychology discussed above in section 2. The OER Research Hub, on which the LACE Evidence Hub was based, found that '[w]ith over a decade's investment in OER there remains surprisingly little reliable empirical research on OER impact' [26]. It seems likely that other areas of TEL research would show the same pattern.

The presence of a large quantity of more qualitative research, theoretical argument, and policy discussion in the learning analytics literature is by no means a weakness. However, as a field founded on the idea of an increase in access to educational data, it is disappointing that there remains so little top-quality quantitative research that demonstrably helps us to achieve improvements in learning and teaching.

The state of the learning analytics literature is in marked contrast to that of the sister field of educational data mining (EDM). The papers in the annual conference and journal of the International Educational Data Mining Society (IEDMS)[10] are overwhelmingly reports of quantitative, empirical work, and there is a growing tradition of making datasets and analysis code available for inspection and re-use, chiefly through the Pittsburgh Science of Learning Center's DataShop.[11] This goes some way to allay many of the concerns currently live in the psychology community discussed in section 2.2. Much of this work concerns self-contained interactive learning material, such as intelligent tutors and simulated lab experiments, rather than the less structured environments studied by most learning analytics researchers.

Efforts have been made in the past to encourage collaboration between the EDM community and the learning analytics community (*e.g.* [4]). It seems that the two could work together to consider the best ways of producing high-quality evidence that benefits learners and teachers. One view might be that EDM

[10] http://www.educationaldatamining.org/

[11] http://www.learnlab.org/technologies/datashop/

provides a natural home for rigorous empirical work, and that learning analytics can employ different standards We disagree: the focus of learning analytics is distinct, with an emphasis on practice, which is question that requires its own strong evidence.

# 5. PROBLEMS WITH THE EVIDENCE

Our analysis of the data in the Evidence Hub highlights important gaps in the evidence that is readily accessible to the learning analytics community, which includes people in a range of different roles, including academics, developers and practitioners.

**Lack of geographical spread:** Our focus has been on the largest area of the Evidence Hub, but the Hub's visualisation tool shows that the majority of widely reported work comes from particular areas of Europe, North America and Australia, with almost no evidence yet emerging from South America, Asia or Africa.

**Gaps in our knowledge:** As noted above, there are particular gaps in the evidence about informal learning, workplace learning and ethical practice, as well as a lack of negative evidence.

**Little evaluation of commercially available tools:** At a time when most learning management systems incorporate some form of learning analytics or data visualisation, we lack evidence that these are having any positive impact on learning and teaching.

**Lack of attention to the learning analytics cycle** [19]: Learning analytics involves the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs [60]. Not enough published work is making it clear how the move will be made from researching the data to optimising the learning; not enough published work is making a connection to the next stage of the learning analytics cycle.

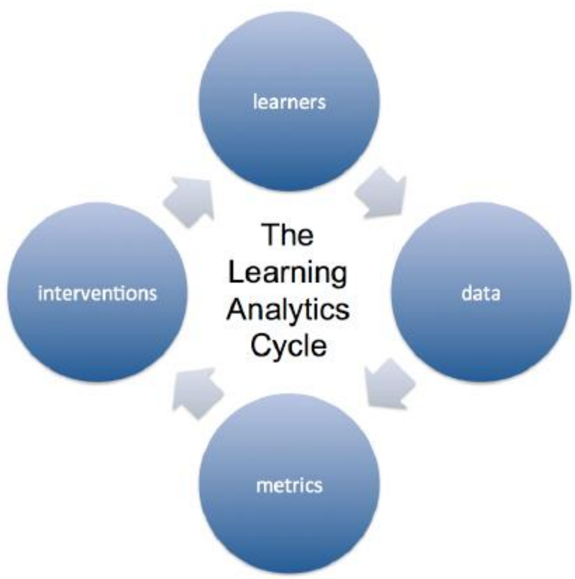


**Figure 2: The Learning Analytics Cycle, from [19]**

**Limited attention to validity, reliability and generalizability:** These are not the only criteria for high quality research that can be used as evidence, but they provide a good starting point. A search for the stems general-, valid- and ethic- in the 22 papers considered here found that 14 referred to validity, eight referred to reliability, seven referred to generalizability and five mentioned none of these. Two papers included 'reliability' as a key word but did not deal with it in the body of the paper. Only four papers [1; 8; 23; 66] included consideration of all three. Almost all papers were based on data from only one institution.

**Limited attention to ethics:** Despite the high levels of discussion of ethical issues in the learning analytics community in recent years, a search for the stem ethic- in the 22 papers considered here showed that only three had explicitly considered ethics. This does not imply that the studies were unethical, simply that the authors did not include any information about how they had dealt with ethical issues when studying hundreds or thousands of learners.

**Sample selection:** The 22 papers here are taken to be evidence of improvement in teaching and learner support. However, relatively few of them include teachers within their sample and in only two cases are there more than 20 teachers within the sample.

**Access to research findings:** The Evidence Hub, which includes brief summaries of papers and full references, is openly accessible. However, the research that sits behind it is often locked away. Of the 22 papers considered here, 18 are LAK papers, sited behind the ACM pay wall ($15 USD for each PDF), one is in BJET ($6 USD to rent the paper, $38 USD for the PDF), and one is a Springer book chapter ($29.95 USD for the chapter). In many cases, pre-print versions are available free of charge from institutional repositories, if you know where to look. However, the default position is to store this research behind pay walls that make it inaccessible to the practitioners and developers who could benefit most from the findings.

**Over-representation of LAK conference papers:** Papers presented at the annual LAK conference are a very important part of the evidence base for learning analytics, but they are only one part of that base. The link between the Evidence Hub and the EasyChair submission system used by LAK ensures good coverage of this conference, but it shifts attention from papers published in a wide variety of journals, from the growing body of literature related to education mining, and from the reports by practitioners and developers that do not appear in a conventional academic format.

# 6. LIMITATIONS

The coverage of the Evidence Hub is focused largely on the LAK dataset. This was by deliberate choice, but does mean there are many other key pieces of research that have not yet been considered for inclusion. In particular, examples of the extensive literature on intelligent tutoring systems (ITS) would be a valuable addition to the Hub. While efforts were made to fully brief reviewers for the Hub on the criteria, to ensure consistency, the decisions made were only lightly crosschecked; the resources available precluded inter-rater reliability checks. The detailed analysis of papers in section 4 is the collaborative work of the two authors, again with no inter-rater reliability checks.

# 7. WHAT IS TO BE DONE?

Now that we are aware of these problems, the learning analytics community can act together to solve them and to establish a firm and accessible evidence base.

The actions proposed here provide possible ways of addressing the problems identified in the previous section. However, if these solutions are to be successful, they need community engagement and community buy-in. In order to start this process, the authors have submitted a workshop proposal to LAK17 that will bring people together from different sectors to discuss and develop the suggestions proposed here. This will be followed by an opportunity at the LAK17 poster session for the entire LAK community to engage with these ideas.

**Evidence Hub**

The LACE special interest group (SIG) of SoLAR now manages the Evidence Hub. The SIG could work to:

- Publicise the Hub and promote engagement – particularly from countries and sectors that have provided little or no evidence to date; pro-active measures may help.
- Identify gaps in the current evidence on a regular basis and share these with the community

**LAK Conference**

The LAK conference committee changes each year, but a set of guidelines could be developed for use or amendment annually.

- Consider how the paper review process could be used to address the problem with evidence. For example, reviewers could be asked to check that all papers either make reference to generalizability, validity, reliability and ethics, or make it clear why this is not appropriate.
- Consider prioritising areas where there are gaps in the evidence in the call for papers.
- Consider how LAK conference papers can be made more accessible to those without access to academic libraries. For example, authors could be asked to supply a separate non-technical summary, and these summaries could be openly accessible.
- Consider measures to strengthen the effectiveness of statistical scrutiny in the reviewing process, while simultaneously encouraging the submission of empirical studies with robust experimental designs.
- Consider requiring authors to specify when they submit a paper how this work fits into the Learning Analytics Cycle (Figure 2), and how it will be connected with the next stage in the cycle.
- Consider more effective ways of sharing expertise with the EDM community.
- Review best practices from fields such as clinical research and psychology that are more advanced in their use of evidence, (e.g. the International Committee of Medical Journal Editors' Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly work in Medical Journals [12]).

**LAK Doctoral Consortium and PhD Supervisors**

- Work to develop and share good practice, and to establish expectations about the quality of evidence
- Help doctoral students to develop research questions and studies that fill significant gaps and fit into the Learning Analytics Cycle (Figure 2).

**Researchers**

- When submitting grant applications, consider how the planned research could be used to fill significant gaps in the existing evidence.
- Consider pathways to impact carefully. How can findings be shared with practitioners who do not read research papers?

**Developers**

- Share work on evaluating tools via the Evidence Hub.

**Journal of Learning Analytics**

- Consider the steps suggested for the LAK conference.
- Where there is a significant body of work available, ask the team or teams responsible to produce an overview paper that brings together the main evidence.

**Society for Learning Analytics Research (SoLAR)**

- Consider making pen drives of past LAK proceedings available to all paid-up SoLAR members, thus providing an access route for non-academics.
- Continue work to engage people from different countries and different sectors.
- Coordinate work across institutions. For example, evaluation of the learning analytics offered by major learning management systems could be carried out at different institutions using the same research design.
- Consider the feasibility and desirability of encouraging pre-registration of empirical studies.

All members of the learning analytics community are encouraged to submit evidence to the Hub. This will make the Hub's coverage more comprehensive, making it easier to identify and then fill the gaps in the evidence.

# 8. CONCLUSIONS

Learning analytics as a field is not immune from the challenges facing empirical research in other disciplines, notably medicine and psychology. These challenges arise from powerful pressures that are far beyond the scope of individual researchers to address, no matter how well-intentioned and well-informed statistically. The nature of the topic area makes it hard to carry out rigorous quantitative research, and rigorous qualitative research is also required to yield not only actionable insights, but also action that improves learning. To validate the field, we must have evidence about whether learning analytics does improve learning and teaching in practice. As a field with an abundance of data, learning analytics should be well placed to produce such evidence. This paper's exploration of the evidence we have to date shows clearly that there is considerable scope for improving the evidence base for learning analytics. We believe that doing so is a scientific and moral imperative. We have set out some suggestions for how we can move forward as a community, and look forward to being part of that work.

## 8.1 Acknowledgement

The European Commission Seventh Framework Programme, grant number 619424, funded the LACE project, which was responsible for developing the Evidence Hub.

# 9. REFERENCES

[1] ALLEN, L.K., MILLS, C., JACOVINA, M.E., CROSSLEY, S., D'MELLO, S., and MCNAMARA, D.S., 2016. Investigating boredom and engagement during writing using multiple sources of information: the essay, the writer, and keystrokes. In *LAK16* ACM, 114-123.

[2] ALLEN, R., 2013. Evidence-based practice: why number-crunching tells only part of the story. Blog post: https://ioelondonblog.wordpress.com/2013/03/14/evidence-based-practice-why-number-crunching-tells-only-part-of-the-story/. In *IOE London Blog*.

[3] ARNOLD, K.E. and PISTILLI, M., 2012. Course Signals at Purdue: Using Learning Analytics To Increase Student Success. In *LAK12* ACM, 267-270.

[12] http://www.icmje.org/recommendations/

[4] BAKER, R.S., DUVAL, E., STAMPER, J., WILEY, D., and BUCKINGHAM SHUM, S., 2012. Educational data mining meets learning analytics. In *LAK12* ACM, 20-21.

[5] BAKHARIA, A., CORRIN, L., DE BARBA, P., KENNEDY, G., GAŠEVIĆ, D., MULDER, R., WILLIAMS, D., DAWSON, S., and LOCKYER, L., 2016. A conceptual framework linking learning design with learning analytics. In *LAK16* ACM, 329-338.

[6] BOS, N. and BRAND-GRUWEL, S., 2016. Student differences in regulation strategies and their use of learning resources: implications for educational design. In *LAK16* ACM, 344-353.

[7] BROWN, M.G., DEMONBRUN, R.M., LONN, S., AGUILAR, S.J., and TEASLEY, S.D., 2016. What and when: the role of course type and timing in students' academic performance. In *LAK16* ACM, 459-468.

[8] BUCKINGHAM SHUM, S., SÁNDOR, Á., GOLDSMITH, R., WANG, X., BASS, R., and MCWILLIAMS, M., 2016. Reflecting on reflective writing analytics: Assessment challenges and iterative evaluation of a prototype tool. In *LAK16* ACM, 213-222.

[9] CAMPBELL, J.P., 2007. Utilizing Student Data within the Course Management System To Determine Undergraduate Student Academic Success: An Exploratory Study, PhD thesis, Purdue University, available: http://docs.lib.purdue.edu/dissertations/AAI3287222/.

[10] CAMPBELL, J.P., DEBLOIS, P.B., and OBLINGER, D.G., 2007. Academic analytics: a new tool for a new era. *Educause Review* 42, 4 (July/August), 40-57.

[11] CAMPBELL, J.P. and OBLINGER, D.G., 2007. *Academic Analytics.* Educause. http://net.educause.edu/ir/library/pdf/PUB6101.pdf

[12] CAMPUS TECHNOLOGY, 2006. Data mining for academic success. In *Campus Technology (21 May 2006)*.

[13] CARNOY, M., 2015. *International Test Score Comparisons and Educational Policy: a Review of the Critiques.* National Education Policy Center.

[14] CAULFIELD, M., 2012. Course Signals and Analytics. Blog post: http://hapgood.us/2012/08/24/course-signals-and-analytics/. In *Hapgood*.

[15] CAULFIELD, M., 2013. A Simple, Less Mathematical Way To Understand the Course Signals Issue. Blog post: http://hapgood.us/2013/09/26/a-simple-less-mathematical-way-to-understand-the-course-signals-issue/ In *Hapgood*.

[16] CHARLEER, S., KLERKX, J., and DUVAL, E., 2014. Learning dashboards. *Journal of Learning Analytics* 1, 3, 199-202.

[17] CHATTERJI, M., 2013. Global forces and educational assessment – a foreword on why we need an international dialogue on validity and test use. In *Validity and Test Use, An International Dialogue on Educational Assessment, Accountability and Equity*, M. Chatterji Ed. Emerald, Bingley, UK.

[18] CHEN, Y., YU, B., ZHANG, X., and YU, Y., 2016. Topic modeling for evaluating students' reflective writing: a case study of pre-service teachers' journals. In *LAK16* ACM, 1-5.

[19] CLOW, D., 2012. The learning analytics cycle: closing the loop effectively. In *LAK12* ACM, 134-138.

[20] CLOW, D., 2013. Looking harder at Course Signals (13 November 2013). Blog post: https://dougclow.org/2013/11/13/looking-harder-at-course-signals/. In *Doug Clow's Imaginatively-Titled Blog*.

[21] CLOW, D., FERGUSON, R., MACFADYEN, L., PRINSLOO, P., and SLADE, S., 2016. LAK Failathon. In *LAK16* ACM, 509-511.

[22] COOPER, M., FERGUSON, R., and WOLFF, A., 2016. What can analytics contribute to accessibility in e-learning systems and to disabled students' learning? In *LAK16* ACM, 99-103.

[23] DAWSON, S., 2009. 'Seeing' the learning community: an exploration of the development of a resource for monitoring online student networking. *British Journal of Educational Technology* 41, 5, 736-752.

[24] DAWSON, S., MCWILLIAM, E., and TAN, J.P.-L., 2008. Teaching smarter: How mining ICT data can inform and improve learning and teaching practice. In *ascilite 2008*, Melbourne, Australia (30 Nov-3 December).

[25] DE LIDDO, A., BUCKINGHAM SHUM, S., MCANDREW, P., and FARROW, R., 2012. The Open Education Evidence Hub: a collective intelligence tool for evidence based policy. In *Proceedings of the Joint OER12 and OpenCourseWare Consortium Global 2012 Conference* (Cambridge, UK, 16-18 April 2012).

[26] DE LOS ARCOS, B., FARROW, R., PERRYMAN, L.-A., PITT, R., and WELLER, M., 2014. *OER Evidence Report 2013-2014* OER Research Hub. http://oro.open.ac.uk/41866/.

[27] ESSA, A., 2013. Can We Improve Retention Rates by Giving Students Chocolates? Blog post: http://alfredessa.com/2013/10/can-we-improve-retention-rates-by-giving-students-chocolates/. In *alfredessa.com*.

[28] FELDSTEIN, M., 2013. Purdue University Has an Ethics Problem (25 November 2013). Blog post: http://mfeldstein.com/purdue-university-ethics-problem/. In *e-Literate*.

[29] GELMAN, A. and LOKEN, E., 2013. The garden of forking paths: Why multiple comparisons can be a problem, even when there is no 'fishing expedition' or 'p-hacking' and the research hypothesis was posited ahead of time. http://www.stat.columbia.edu/~gelman/research/unpublished/p_hacking.pdf

[30] GINSBURG, A. and SMITH, M.S., 2016. *Do Randomized Controlled Trials Meet the "Gold Standard"?* Blog post: http://www.aei.org/publication/do-randomized-controlled-trials-meet-the-gold-standard/. In *American Enterprise Institute*.

[31] GOLDACRE, B., 2013. *Building Evidence into Education.* Department for Education, UK. https://http://www.gov.uk/government/news/building-evidence-into-education.

[32] GOVAERTS, S., VERBERT, K., and DUVAL, E., 2011. Evaluating the student activity meter: two case studies. In *International Conference on Web-Based Learning* Springer, 188-197.

[33] GREENHALGH, T., HOWICK, J., and MASKREY, N., 2014. Evidence based medicine: a movement in crisis? *BMJ* 2014;348:g3725

[34] HARRISON, S., VILLANO, R., LYNCH, G., and CHEN, G., 2016. Measuring financial implications of an early alert system. In *LAK16* ACM, 241-248.

[35] HENRICH, J., HEINE, S.J., and NORENZAYAN, A., 2010. The weirdest people in the world? *Behavioral and Brain Sciences* 33, 2-3, 61-83.
[36] HOWARD-JONES, P.A., 2014. Neuroscience and education: myths and messages. *Nature Reviews Neuroscience* 15, 12, 817-824.
[37] IOANNIDIS, J.P., 2005. Why most published research findings are false. *PLoS Medicine* 2, 8, e124.
[38] IOANNIDIS, J.P.A., 2016. Why Most Clinical Research Is Not Useful. *PLoS Medicine* 13, 6, e1002049.
[39] IZAWA, M.R., FRENCH, M.D., and HEDGE, A. Shining new light on the Hawthorne illumination experiments. *Human Factors* 53, 5, 528-547.
[40] JOKSIMOVIĆ, S., MANATAKI, A., GAŠEVIĆ, D., DAWSON, S., KOVANOVIĆ, V., and DE KEREKI, I.F., 2016. Translating network position into performance: importance of centrality in different network configurations. In *LAK16* ACM, 314-323.
[41] KARKALAS, S. and MAVRIKIS, M., 2016. Towards analytics for educational interactive e-books: the case of the reflective designer analytics platform (RDAP). In *LAK16* ACM, 143-147.
[42] KIRKWOOD, A. and PRICE, L., 2015. Achieving improved quality and validity: reframing research and evaluation of learning technologies. *European Journal of Open, Distance and E-learning* 18, 1, 102-115.
[43] KLENOWSKI, V., 2015. Questioning the validity of the multiple uses of NAPLAN data. In *National Testing in Schools: An Australian Assessment*, B. Lingard Ed. Routledge, 44-56.
[44] LAURÍA, E.J.M., MOODY, E.W., JAYAPRAKASH, S.M., JONNALAGADDA, N., and BARON, J.D., 2013. Open academic analytics initiative: initial research findings. In *LAK13* ACM, 150-154.
[45] MATHEWSON, T.G., 2015. Analytics programs show 'remarkable' results — and it's only the beginning. Blog post: http://www.educationdive.com/news/analytics-programs-show-remarkable-results-and-its-only-the-beginning/404266/. In *Education Dive*.
[46] MOSTAFAVI, B. and BARNES, T., 2016. Data-driven proficiency profiling: proof of concept. In *LAK16* ACM, 324-328.
[47] MUSLIM, A., CHATTI, M.A., MAHAPATRA, T., and SCHROEDER, U., 2016. A rule-based indicator definition tool for personalized learning analytics. In *LAK16* ACM, 264-273.
[48] OPEN SCIENCE COLLABORATION, 2015. Estimating the reproducibility of psychological science. (28 August). *Science* 349, 6251.
[49] OSTER, M., LONN, S., PISTILLI, M.D., and BROWN, M.G., 2016. The learning analytics readiness instrument. In *LAK16* ACM, 173-182.
[50] PAPAMITSIOU, Z., KARAPISTOLI, E., and ECONOMIDES, A.A., 2016. Applying classification techniques on temporal trace data for shaping student behavior models. In *LAK16* ACM, 299-303.
[51] PETTICREW, M. and ROBERTS, H., 2003. Evidence, hierarchies, and typologies: horses for courses. *Journal of Epidemiology and Community Health* 57, 7, 527-529.
[52] PISTILLI, M.D. and ARNOLD, K.E., 2010. Purdue Signals: Mining real-time academic data to enhance student success. *About Campus: Enriching the Student Learning Experience* 15, 3, 22-24.
[53] [PSATY, B.M., WEISS, N.S., FURBERG, C.D., KOEPSELL, T.D., SISCOVICK, D.S., ROSENDAAL, F.R., SMITH, N.L., HECKBERT, S.R., KAPLAN, R.C., LIN, D., and FLEMING, T.R., 1999. Surrogate end points, health outcomes, and the drug-approval process for the treatment of risk factors for cardiovascular disease. *Journal of the American Medical Association* 282, 8, 786-790.
[54] RIENTIES, B. and TOETENEL, L., 2016. The impact of 151 learning designs on student satisfaction and performance: social learning (analytics) matters. In *LAK16* ACM, 339-343.
[55] RIENTIES, B., TOETENEL, L., and BRYAN, A., 2015. Scaling up learning design: impact of learning design activities on LMS behavior and performance. In *LAK15* ACM, 315-319.
[56] ROBINSON, C., YEOMANS, M., REICH, J., HULLEMAN, C., and GEHLBACH, H., 2016. Forecasting student achievement in MOOCs with natural language processing. In *LAK16* ACM, 383-387.
[57] RUST, J. and GOLOMBOK, S., 2009. *Modern Psychometrics, Third Edition*. New York, London.
[58] SACKETT, D.L., 1997. Evidence-based medicine. *Seminars in Perinatology* 21, 1, 3-5.
[59] SCHEFFEL, M., DRACHSLER, H., STOYANOV, S., and SPECHT, M., 2014. Quality indicators for learning analytics. *Educational Technology & Society* 17, 4, 117-132.
[60] SIEMENS, G., GAŠEVIĆ, D., HAYTHORNTHWAITE, C., DAWSON, S., BUCKINGHAM SHUM, S., FERGUSON, R., DUVAL, E., VERBERT, K., and BAKER, R.S.J.D., 2011. *Open Learning Analytics: An Integrated and Modularized Platform (Concept Paper).* SOLAR.
[61] SIMMONS, J.P., NELSON, L.D., and SIMONSOHN, U., 2011. False-positive psychology undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science* 22, 11, 1359-1366.
[62] SMALDINO, P.E. and MCELREATH, R., 2016. The natural selection of bad science. *Royal Society Open Science* arXiv preprint arXiv:1605.09511.
[63] SUTHERS, D. and VERBERT, K., 2013. Learning Analytics as a 'Middle Space'. In *LAK13* ACM, 1-4.
[64] TARAGHI, B., SARANTI, A., LEGENSTEIN, R., and EBNER, M., 2016. Bayesian modelling of student misconceptions in the one-digit multiplication with probabilistic programming. In *LAK16* ACM, 449-453.
[65] TEMPELAAR, D.T., RIENTIES, B., and GIESBERS, B., 2015. Stability and sensitivity of learning analytics based prediction models. In *7th International conference on Computer Supported Education*, Lisbon, Portugal, 156-166.
[66] WOLFF, A., ZDRAHAL, Z., NIKOLOV, A., and PANTUCEK, M., 2013. Improving retention: predicting at-risk students by analysing clicking behaviour in a virtual learning environment. In *LAK13* ACM, 145-149.

[67]