

Learning analytics in a time of an insatiable thirst for data and evidence: A provocation

PAUL PRINSLOO, University of South Africa, UNISA, South Africa



International Seminar on “Evidence-based research: methodological approaches and practical outcomes. Insights for Online Education”

OPEN SESSION - Will the future of Higher Education be evidence-based? - Barcelona, 22 November 2017

My purpose in my presentation will be to destabilize some of our assumptions about data and evidence. I am not going to change the world and I accept that. I just want to slow the conversation down to say, “Let us reconsider some of our assumptions about data and some of our assumptions about evidence and how it plays out in learning when we evaluate students’ learning.”

The title of this presentation had me thinking. Will the future of higher education be evidence-based? I was thinking maybe there are some more important questions to ask, with a few apologies to the host. I think it is almost a non-question. I would rather say: why is there a need for evidence? Is it about efficiency? Are we worried? Do we want to increase efficiency? Do we want to increase transparency about how we spend the money and how we allocate resources? Is it about accountability?

Just as important a question: who will define what counts as evidence? And what is not regarded as evidence? I think while we say we want evidence and we have the data to back up the evidence, we must also consider what will we not consider as evidence in a particular situation. And there we cannot underestimate the role of gatekeepers. The specialists, the experts, the consultants, the organizations that ensure quality. They have a very specific understanding of evidence, what it is and what it is not. Those who formulate criteria understand this. And then the whole tension between quantitative and qualitative: in some circles, qualitative data is just not regarded as evidence.

Who will verify the evidence? That is one of my concerns. Who will verify the evidence as valid and appropriate for the purpose for and the context in which it was collected? Do we consider that when we talk about evidence? Who will use the evidence and for what purpose? We produce reports and file them and I am not sure we use them. And what is the relationship between data evidence intelligence that seems to be lacking in many governments? Knowledge and

wisdom is even lacking more.

And finally, how do these questions impact the collection, the analysis and the use of student data? So when we look for evidence of learning, how do all these questions impact on why we collect the data, how we use the data and what we make of the data?

If we map the student's learning journey and if we talk about the data that we will collect as students engage in their learning journey, we have a student and we have a learning journey and hopefully we have success or failure. More often, we have early departure. So we collect data about the student at the moment they engage with a higher education institution or with the discipline, and then we collect data throughout the learning journey, whether it is a formative or summative assessment. We look at the number of downloads, we see how many times they logged in, so we collect evidence as they go on with their learning journey and we try to say what is happening in this journey. And we call that descriptive analytics. We try to describe what is happening. But it is not only descriptive; we also want to know why is it happening. And I am not always sure our research and our evidence gets to the why. We are good at predicting patterns. You say, "Oh, there's a pattern that most students that are female and 30 years old with two children and employed do better than males." But why? What is going on there? What will happen? That is predictive. That we will say that if we understand the descriptive and the diagnostic, can we predict that students like them will then act in a particular way?

I want you to notice that with every level of analysis there are different ethical issues. And I would almost have a gut feeling that the responsibility to seriously consider what we regard as data increases as we go down. The ethics and the descriptive analytics is almost for me not as huge as when we start to predict student throughput based on certain behaviour or characteristics. And what scares me most is how can we make it happen? If students of a particular age group, gender, or race act like this, can we prevent them from registering for a particular course? And say, "Based on our evidence, students like you won't make this, so you cannot register." So you see how this plays out.

So all our descriptions, our diagnostics, predictive, and prescriptive analytics feed back into this learning journey. My question is: what data do we need to describe, understand, predict and prescribe the learning journey? And as you see, it's not linear. The students go up and down. And then what data do we already have, in which formats, for what purposes, where are the data stored and governed by whom? So when we talk about data, it seems that we want to have more, but we forget what we already have. The library has data, registration has data, student administration has data, student support has data. Somehow, we want more data. And my first question is: what data do we already have, for what purposes was that data collected, where is it stored, what is the quality of that data, and governed by whom? Who has access to that data? Only then get to the point that we say what data don't we have, and that we need, in order to describe, understand, predict and prescribe? What are our assumptions about learning and data as evidence? And that's an interesting discourse, one to remark later.

This question we never consider: what data do students need to make better informed decisions about the choices they have or they don't have? So we have a lot of data. But I have a suspicion we also need to ask students: "What data would you like to have access to that will help you to make better decisions?"

Let's reconsider learning analytics. So learning analytics as a process actually defines what we consider learning to be and what learning is not. So it is a huge responsibility to say: when do we know a student learns? The number of logins? The number of engagements, whether they are emotionally engaged, behavioural engagement, cognitive engagement? How do we define, what do we measure, how do we define progression and disengagement? Learning analytics decides samples, what data to collect and when. It makes a difference when you collect data three weeks into the semester or when you collect data five weeks into the semester, before the assignment or after the assignment. So these things play a role. What to use as proxies and what data to ignore? Increasingly, higher education uses a student's address, their home address, to make predictions on the probability of success. How does it work? If the student's address is in a poor neighbourhood with a low socioeconomic income, of a particular race, then we use the home address as a predictor to say: with these characteristics, we do not expect you to pass this course. A home address becomes then a powerful tool that the administration used to predict what is the likelihood of this student paying for their fees. That's dramatic. That raises a number of ethical issues.

What data do we ignore, most probably because we are more comfortable with not taking account of that data? Who will interpret that data? Google does not appoint female analysts. Does a gender and a race and a background and an epistemology and a worldview of the analysts make a difference in how they interpret data? Yes, it does. There is enough evidence for that. So who will and how do we interpret analysis and findings? Learning analytics will also decide what findings to share with whom, as well as choosing methods of dissemination. So whether it is a report or it is a dashboard, whether it's early alert systems.

So our understanding of learning analytics as a process of collecting evidence and measuring success and efficiency is shaped by how we understand and describe students and their learning. If we talk about students as customers, we will collect a certain type of information. If we see them as learners, we will look for different types of information. And this was this morning on the online journal *EdSurge News*: What can online education learn from precision medicine? And they talk about patients as being broken and sick and they say higher education should look at students as broken and sick, and in need of healing. They come into the system broken and higher education must fix them. I nearly collapsed and I said, "What?" What happens when you think of students as broken and higher education must fix them? I just said, "I cannot repeat that."

So how will we know that learning analytics is having a positive impact? That is what evidence hopefully is about. Is impact more successful or satisfied students? I have a suspicion that satisfaction is not necessarily successful or the other way around. My students in my course that I teach online are probably very dissatisfied with the marks I give them, but they are successful. I make sure

they pass. Is it about more effective or more appropriate teaching? Is it about better allocation and utilisation of resources? How will we know that learning analytics and all this evidence will have an impact? This was published earlier this year by Rebecca Ferguson and Doug Clow from the Open University. They asked: "Where is this evidence? Learning analytics has been around since 2011, and with all due respect, we don't have enough evidence that it really makes a difference."

Since 2011 we have been learning analytics and have been collecting student data, we analyse, and we go to nice locations like Barcelona to make presentations. We do not have evidence that it really impacts on teaching and learning. I want to propose that we understand the evidence of impact or the lack thereof as entangled in our assumptions about evidence, data, and the messy processes informing implementation. Let me explain. I consider evidence as contested, increasingly. It is political. It is incomplete and it is fragile. And don't throw me out yet. I will explain. The antiretroviral program in South Africa, that is medication for people that have tested positive for HIV/AIDS, you will see from 2009 to 2016 there has been a steady increase of the amount of people on the ARV programme, with huge success. But what happened before 2009? HIV/AIDS and antiretroviral treatment was there since the late 1990s. We had a president, and in his reign, from 1999 to 2008, he disputed the evidence of most scientists in the world that said HIV/AIDS is caused by a virus. The president of a country refused to roll out ARV treatment and 300,000 people, and that is an estimate, died.

When we talk about evidence, we should look at evidence at the intersection of politics and policy. Example two: and this is from the delightful book by Loren Graham. Problem: Russia, 20th century. We need a more sustainable water supply. Stalin: "What is the problem? Do some research. Propose an evidence-based plan." The engineer, Peter Palchinsky, says, "Smaller dams will be more effective." Stalin says, "I want to build the highest, largest hydroelectric dam." Stalin won. Peter Palchinsky was executed. Not because he wanted to build smaller dams, but he was executed three years later. And the outcome of this project was massive costs with 100,000 people displaced.

Example three: evidence at the intersection between data and zeitgeist. This is from 2017, November 16th. That is four days ago. "We have compiled the list of 30 scientific studies that show a link between vaccines and autism, disproving the myth that no official research papers exist to support what alternative doctors have been saying for years." When I look at who publishes it, it is *Newsbunch*. When I look further, it is a fake news site. But thousands of people have stopped vaccinating their children because they found 30 scientific articles that disprove the link. And I said, almost like that first link, "What?"

Evidence is really contested. We cannot take evidence as speaking for itself. It is political, it is incomplete. It is really a fragile concept, despite our standards and our criteria and our Excel spreadsheets. So how does learning analytics define, produce and use evidence? 2012: learning analytics is the new black. 2015: a revolution. Another one. Oh sorry, I should not say that in Catalonia. "Higher education and the revolution of learning analytics." There is no yellow and red on this slide. "Stanford calls for responsible use of student data in

higher education.” That is 2016. “From high school to Harvard, students urge for clarity on privacy rights.” Some students say, “I don’t want you to get all this data from me. I am concerned how you use this data. Even if it’s for my benefit, I want a say in this data.” Then of course, the revolution just continues. Tracking every student’s move. We want to put trackers on students so we can see when they go to the restaurant or the toilet or they do not move.

So eight provocations for thinking about evidence in a half way. How do we think about evidence in a world saturated with data? This is from Michael Patton, that says with so much evidence and information around us, the main challenge is how to select data. Data is not neutral. It is there. It is everywhere. Our technological capacity for gathering and computerizing information has far exceeded our ability to process and make sense of it. How do we think about evidence in a world of fake news, myths, alternative facts and botched truths? We have an American president that has his alternative truths. There are neuro-hits and neuro-myths. This is from *Nature* a few days ago: “Neuroscience and education: myths and messages.” This is my favourite hobbyhorse: “All you need to know about learning styles” myth in two minutes. It is a myth that would not die despite all the evidence that says that there is no such things. I think three quarters of my institution’s faculty believe in learning styles. I wanted to say, “What are you reading?”

How do we think about evidence in a world where knowing is distorted and manipulated to create biased and distorted findings? The vaccination is an example. This is the best-known example: Bush knew Saddam had no weapons of mass destruction and they went to war. How do you think about evidence in a world where what we know is increasingly determined by algorithms and automated agents on Google, Amazon and Facebook? This is from 30th October, André Staltz, that says Google, Facebook and Amazon determine 70% of the traffic on your internet. They decide what you read. They decide what our students read. This is my favourite author posted 10th of November, Pasquale, Frank Pasquale, that says: instead of a wealth of networks, it has given way to a black box society where trolls, bots, and even foreign governments maraud to distort the information environment on Twitter, Facebook. I published this article where I say: but, and if we start to use algorithms to use student data, we are creating a monster like Dr Frankenstein. As we try to get away from the monster of Dr Frankenstein, we meet Kafka. We never get out. We must use the data. Then it can become a monster. Therefore, we are in this dance with destiny.

How do we think about evidence in a world where knowing does not mean we will act? I know I must go to the gym, but I do not. This is more serious than going to the gym. This is a report from lessons from Rwanda. The refusal of international agencies and world leaders to take seriously and use the data they were given resulted in a genocide where millions of people were killed. How do we engage with evidence in a world where knowing does not mean we have the resources or capacity to respond effectively and appropriately? Increasingly, in a resource-constrained higher-education sector, we classify students as: almost dead, let’s not spend money on them; if we spend a lot of money on these, they may get back onto the battlefield; but if we spend money on these students, with a little help they can do better. We do not have unlimited resources. Therefore, we just let some students die because we do not have the resources. So

evidence, I would like to propose, is really, really a fragile concept and because we have the evidence it does not mean we have the resources to use it.

Lastly, how do we collect evidence in a world where evidence of efficiency and effectiveness does not mean that the intervention was appropriate, moral, or ethical? The gas chambers in Auschwitz were very effective. But it was immoral. Zygmunt Bauman wrote a book on how modernity and the industrialization process actually resulted in the Holocaust. And how do we present evidence in a world where evidence will only be recognized and considered if the problem is recognized as serious and requiring a new solution, when the policy community develops a financially and technically workable solution, and when political leaders find it advantageous to approve it? We have the evidence, we have the solution, and it may never happen because political leaders will only use it when it is to their advantage.

This is a wonderful book that came out in 2016. I only bought it last week. Alice Dreger: *Galileo's middle finger: Heretics, Activists and One Scholar's Search for Justice*. And she says the pursuit of evidence is probably the most pressing moral imperative of our time.

I am wrapping up, so let's reconsider learning analytics. Gert Biesta says evidence-based education favours a technocratic model, where it is more important to consider something as effective, and not consider whether it is educationally desirable. This is Gert Biesta again that says it is entangled at the intersection of the impatience of policy makers to find quick solutions, stakeholders on the ground, and specialists. These are his two articles and I would really propose that you engage with [them](#): "Why 'what works' won't work: Evidence-based practice and the democratic deficit in educational research," this is from 2007, and he followed it up in 2010 with "Why what works still won't work." We are so concerned with what works and he says that is not the right question. Let's reconsider data. And just shortly, we have access to more volume, velocity, variety of student data than ever before, and it allows us to expand the scope of our research, but also to make inferences unthinkable a number of years ago. So we need to seriously consider our own assumptions and epistemologies. And this is from Nate Silver, that says noisy systems with underdeveloped theory are very dangerous. So we may mistake the noise for the signal. And we know what the management of a university does, when they think they have a signal.

Data is not neutral. It is not raw. It is not objective. It is not pre-analytic. Data are framed technically, economically, ethically, temporally, spatially, philosophically. This is scary: data do not exist independently of the ideas, instruments, practices, contexts and knowledge used to generate processes and analyse them. Data do not speak for themselves. That is a claim by Miles Schoenberg and Coupier. It is not enough to know that people are doing things without understanding why. Miles Schoenberg said, "Big data will be the truth. We are getting to the point that all our questions will be answered." And I just question that. Data sets represent cultural, moral, instrumental process. More data is not necessary. Better data. The sheer size of the analysis does not eschew the limitations of subjectivity. And I love this from [Papacharissi](#): the unbearable

lightness of information. It is fragile.

So I ask: what happens when you collect, analyse and use the bleeps and the logins, the downloads, the posts, outside of the institutional learning management system, as a full and the only narrative of what is happening in our students' lives, of their aspirations, and their learning? What happens if we think that's the truth and we make decisions on their number of logins? What are the implications when we use their data points to describe, diagnose, predict and prescribe their learning journeys without ever asking them what these data points mean to them? And what data would matter to them to allow them to make better choices? Do we ever ask students, "We see you haven't logged on this week. Do you want to talk about it?" And not only use it as a predictive value for their probability of failing.

How does evidence look and function when we consider education as an open and recursive system, worlds apart from research environments where we can control variables? And I have, I really doubt randomized control trials in education. I just think they are immoral. How does evidence look when we consider students' success? And this is very convoluted, I do apologize. When we consider student success as complex, non-linear, and the result of intersecting, often interdependent and mutually constitutive variables in the nexus, in the middle between students' life worlds, their capital and context, institutional assumptions, epistemologies and inefficiencies and macro societal shifts?

Can I just explain it now? And I am not going to spend time on this. You have the student, this agent, where they were born, whether they are black, female; have children out of wedlock, whether they are married... You have the student, you have the institution here, and then when these two meet, you have choice/admission, learning activities, course success, graduation, hopefully employment, and success. And then you have shaping conditions on both sides. Suddenly there is a postal strike. Suddenly the Internet goes down. It is outside of the control of the student. It is outside of the control of the institution. And it impacts on the journey. Suddenly the student loses his job. Suddenly she falls pregnant. Suddenly they test positive for HIV. It is a life-changing condition outside the control and it impacts on this journey. And in the light of this complex system, you want to measure the number of logins? And not account for the complexity of this? I have got the reference and you can follow it up.

Some pointers to finish: I see learning analytics as moral practice. Just because you can collect the data doesn't mean you have to. And even if it may be legal it may not be ethical. Learning analytics should not only focus on what is effective but aim to provide relevant pointers to decide what is appropriate and morally necessary.

Second pointer: if learning analytics' primary aim is to improve students' learning, maybe we should ask: to what extent has our learning analytics become our voice over our students' learning experiences by telling them, "This is what your learning looks like and it's the only narrative that counts?" How can we replace our epistemological arrogance and colonization of student experience with an epistemological shift from knowing to listening, humanity and

respect? It is a major shift. In the one, we will see students as broken and in need of healing and the other one, we will say, "What does this mean?"

What happens if we stop talking about students as sick, broken and dropouts? What happens when we discard the deficit model of student learning to one that focuses on what they have, what they bring to the table? Then we would gather different types of data. What data do we have that if they have access to the data it will allow them to make better informed choices? And what data to they have that if we have that data it will help us to provide more effective and appropriate learning?

But it is also our story, and there's just three slides here. Educational triage: this is from 2014. If we identify students at risk and we do not have the resources to respond, it creates a moral dilemma. We followed it up by this set to say that institutional survival is at stake. And it is the elephant in the learning analytics room, the obligation to act. Very shortly on here: does the contractual obligation between the institution and the student force the institution to act on what they know? There is only one case study in law, where a student committed suicide at an American higher educational institution, and the parents took the institution to court because the institution knew the student had psychological problems. And the court awarded damages to the parents. And that raises a very interesting point: if we know a student is at risk, if we have the evidence and we don't act, we may be taken to court. So be careful whether we want to know more. The political nature of learning analytics, it is a structuring device. It is not neutral. It is informed by current beliefs about what counts as knowledge and learning. It has coloured by assumptions about gender, race, class, capital, literacy, and it is in service of perpetuating existing or new power relations.

The second-to-last pointer: how do we collect and use data when the data represents the results of generations of structural inequalities and injustices? When the home address of a student is outside their control? How do we use learning analytics to break intergenerational cycles of bias, discrimination and exclusion? Consider the difference between correlation and causation in a complex system like that. And I'm not going to say any more about that.

Who oversees learning analytics? If I want to do research on students I must go through an ethical review process that takes three months, twenty signatures, and most probably five committees, but learning analytics engage with student data and they influence student behaviour without any oversight. And I don't say that we should go through another ethical review board process. But who oversees the ethical implications of when we collect and use student data, outside of research with a capital "R"? Higher education cannot afford not to collect and analyse the new student data. We have a contractual duty, we have a fiduciary duty, we have the opportunity and the authority to act and respond. We do not have unlimited resources. We need to be transparent and accountable for what we can and cannot do. And once we know, we cannot unknow knowing; we have a moral duty to respond.

Hence, if learning analytics is contested, political, incomplete and fragile, I propose that we see evidence as an invitation to conversation and not the end of a conversation. The question, therefore, is not: "Will the future of higher

education be evidence-based?”, but “How do we define evidence?” Who defines it and what and whose purpose will it serve when acted upon?