Learning Analytics: A Primer

Learning Analytics:
A Primer

COMMONWEALTH of LEARNING

Learning Analytics: A Primer

The Commonwealth of Learning (COL) is an intergovernmental organisation created by Commonwealth Heads of Government to promote the development and sharing of open learning and distance education knowledge, resources, and technologies.

Authors:
Paul Prinsloo, *University of South Africa*
Sharon Slade, *Earth Trust, United Kingdom*
Mohammad Khalil, *University of Bergen, Norway*

Published by:
Commonwealth of Learning
4710 Kingsway, Suite 2500
Burnaby, British Columbia
Canada V5H 4M2
Telephone: +1 604 775 8200
Fax: +1 604 775 8210
Web: www.col.org
Email: info@col.org

Table of Contents

Introduction

Welcome to *Learning Analytics: A Primer*. As the word “primer” indicates, it is a surface-level or first-layer introduction to the complex world of learning analytics.

In all levels of education, from pre-primary to post-secondary, teachers have a contractual and moral duty to facilitate learning. It does not matter whether you are teaching face-to-face classes in a traditional (residential) setting, or teaching online or via correspondence courses, our aim as teachers is to facilitate learning. But how do we know whether our learners are learning and have reached the required levels of competency or understanding? Through different summative assessment strategies, we get a sense of whether learners have achieved the intended outcomes, but by then, it is most probably too late to intervene for those learners who have not managed to reach the required levels of competency or understanding. Good teaching therefore involves a number of formative assessment strategies throughout the semester or study period, to ascertain to what extent learners are making progress, and whether they are having difficulty with particular concepts. Once a formative assessment opportunity provides evidence that a learner may be struggling or is not coping, we can intervene and address whatever the issue is.

While the above may be all too familiar, we don’t necessarily see the collection of evidence of whether or not learners are learning as something “special” — it is, in many ways, just what good teaching looks like!

This provides us with a useful basis on which to consider learning analytics. Collecting and analysing formative and summative assessment data to make judgements about whether learners need extra support and motivation, or whether they are coping and will achieve the desired outcomes, are largely what learning analytics is all about.

Since its emergence in 2011 as a distinct discipline, research focus, and practice, learning analytics has matured and grown into an established field of inquiry. The definition of learning analytics initially established and still relevant today is: “the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimising learning and the environments in which it occurs.”[1]

If we combine the definition of learning analytics with the scenario of understanding learner progress, and move data generation or teaching to an online environment, we possibly have access to more data than just formative and summative assessment outcomes. For example, when learners log onto a course platform, they leave data trails that provide us with some sense of how they progress through the course, which resources they download (possibly more than once), who they interact with, how often they engage with the content and/or pose questions and so forth. All these data points (see Units 2 and 3) can assist us in understanding their learning journeys and in identifying those learners who may need extra support or stimulation.

Welcome to the world of learning analytics!

Much has been written about learning analytics, and the body of research underpinning the growth of learning analytics is remarkable. However, the field may be inaccessible to teachers, learners and the general public, who might not ordinarily have access to the published research or possess the necessary academic or practical experience to make much sense of the field.

This resource intends to provide a brief introduction to key issues around learning analytics. While we have aimed to generate an accessible introduction to learning analytics, we have not compromised on scientific rigour and good scholarship!

[1] https://www.solaresearch.org/about/what-is-learning-analytics/

Watch Video: https://www.youtube.com/watch?v=DwUv-gFpLyU

Video attribution: "Unit 0: Welcome to Learning Analytics - A Primer" by Commonwealth of Learning is available under CC BY-SA 4.0.

Purpose of the Course

The course has been designed and written with the explicit purpose of assisting teachers to understand and use learning analytics ethically and appropriately in their own context.

Teachers form the central focus of the course. Learning analytics as a research field and practice is, at its core, interdisciplinary and includes theories, methodologies and practices from a range of disciplines, such as education, sociology, computer science, mathematics, data science and psychology, to mention but a few. We have tried to make the text and key concepts as accessible as possible for a broad audience. Some of the units referring to mathematical and/or computer science concepts may be more challenging for those without a background or interest in these concepts. We have attempted, though, to make those materials as accessible and clear as possible.

We foresee that as you work through this course and complete the activities and self-assessment questions in each unit, you will gain not only an informed understanding of learning analytics but also the ability to apply the principles and practices of learning analytics to your own particular context.

Learning Outcomes for the Whole Course

On completion of this course, participants should be able to do the following:

1. Explain learning analytics in the broader evolution of the use of learners' data in teaching and learning environments.
2. Critically examine the claims of objectivity and measurement made in the world of data.
3. Provide a personalised and contextualised definition of learning analytics and distinguish learning analytics from academic and teacher/teaching analytics.
4. Distinguish between different sources of learners' data, and map data they have access to in their own contexts, to inform their teaching and their learners' learning.
5. Explain the different uses of learning analytics and develop a strategy to use learning analytics at a course level in their own contexts.
6. Demonstrate a basic, informed understanding of working with data and the different free software available to assist teachers.
7. Identify ethical issues in learning analytics, take the necessary steps to protect learner privacy, and create a personalised statement of consent for use in their own classrooms/contexts.
8. Create a basic implementation plan for using learning analytics in their own teaching.

9. Appreciate the critical role of a policy framework to guide the implementation of learning analytics.
10. Describe some of the future trends in learning analytics and how these might apply within their own contexts.

Structure of Each Chapter

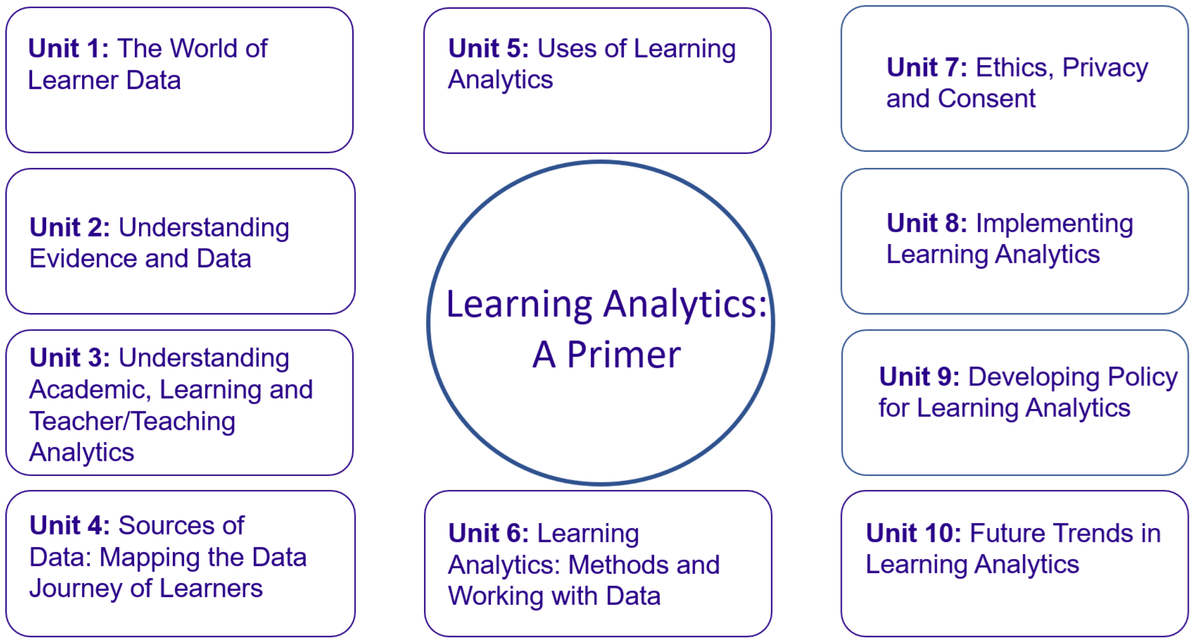
Each of the units is organised in a similar way. The unit starts with a brief introduction to its focus. This is followed by an overview of the planned outcomes. Self-assessment is an integral part of each unit, and you will have two opportunities in each unit to check your understanding of its focus by answering questions. Responses to these questions are provided at the end of each unit.

Each unit makes use of tables, figures and diagrams to break up the text and enhance the transmission of concepts. Some units also include short videos to highlight particular aspects of the topics covered.

In addition to the content, activities and self-assessment questions in every unit, you will find a number of resources and references in the footnotes.

An Overview of the Units

There are ten units in this course, and each explores a specific aspect of learning analytics (see Figure "An overview of the units in this course" below). While you are welcome to access the units in any particular order according to your interests, we have designed this primer around the units being in a particular sequence. For example, the first five units provide an important foundation for the rest of the course. We suggest that you consider doing units one to five before you explore the remaining units.



An overview of the units in this course

Conclusion

We sincerely hope you enjoy working through the units, watching the videos and self-assessing your understanding of learning analytics as you progress through the course. We wish you all the best in your exploration!

UNIT 1

The World of Learner Data

Introduction

Though learning analytics emerged as a new field of inquiry and practice fairly recently (2011), it is important to situate it in historical context — specifically, that educators have always collected, analysed and used learner data for a variety of purposes, such as planning, strategy, assessment, reporting and quality assurance.

Since the earliest recorded human activity, the evidence that individuals had learned something — for example, crafting utensils, finding their way to a destination or finding food — was provided once these tasks were successfully completed. One could not claim, for example, to be a (good) hunter if nothing was caught. The proof of learning has been found in *evidence*. (For an exploration of evidence and data-as-evidence, see Unit 2.)

In this unit, we situate learning analytics in a historical context of formal education where learner data have always been collected, where progress was measured and reported on, and where educational institutions have used these data for a variety of purposes.

Learning analytics is a continuation of this long history — collecting, measuring, analysing and *using* learner data to help learners and teachers understand learners' learning and their progress.

Learning outcomes

After working through this unit, you should be able to:

- explain how education systems have always used learner data for a range of
- purposes, such as planning, operations, budgeting, pedagogical decisions and support

- map your institution's collection and use of learner data — what data are collected and for what purposes, where the data are stored, who has access to the data and under what conditions
- create a case study of one of the courses you teach, and consider what data you use to make informed decisions on issues of pedagogical design, assessment and learner support

Collecting and using learner data: a short historical overview

In the introduction to this unit, we pointed out that evidence or data have always been part of human learning. When education became formalised, various types of evidence began to inform how to arrange classes, how to plan for the next academic year, how to design curricula and outcomes, and, importantly, how to design assessment opportunities to evaluate whether learners have achieved the desired learning outcomes. Before the digital era, learners completed paper-based admission forms, and these forms provided school administrators and teachers with useful information about learners, such as their prior learning experiences, demographic details, and so forth. Learners were allocated to classes and to teachers who accompanied the allocated learners for the duration of the academic year. While the curriculum and the assessment strategies may have been predetermined (e.g., in the preceding academic year), teachers could adapt the curriculum and their pedagogical strategies based on observing learners — identifying which had trouble with some concepts, and which became bored waiting for others to catch up.

Observational data were, however, not usually enough to provide evidence of learning, so teachers designed formative assessments to determine learners' progress. Marking these assessments served two important purposes, the first of which was to provide learners with feedback on their understanding and competencies in a specific area. These formative assessment opportunities also provided teachers with important information on where learners misunderstood concepts, which learners needed additional resources or extra attention, and which learners needed additional stimulation.

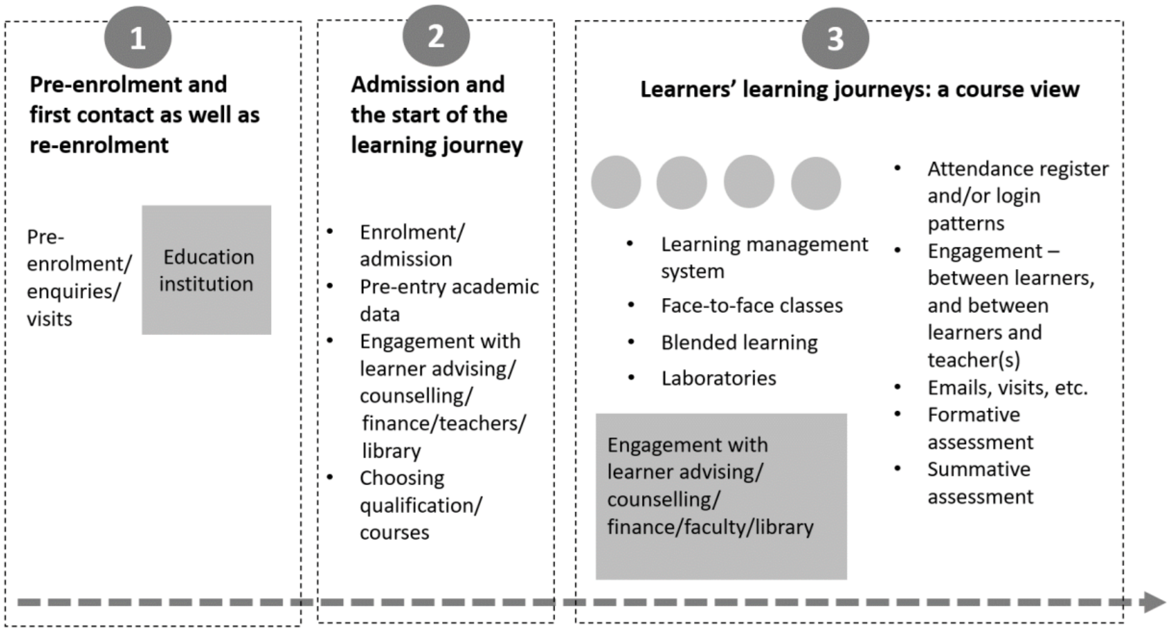
At the end of the academic period, teachers also had a responsibility to ensure that learners had sufficiently attained the intended outcomes to proceed to the next level. The level of attainment also served a dual purpose: as feedback to learners but also to the school administration to support them in placing learners at the right level for the following academic period.

Thinking in terms of the above historical background, we can see that data about learners has been part and parcel of education for a very long time. These data have served a number of purposes: planning, teaching, assessment and quality assurance.

Learner data: an institutional perspective

It does not matter to what extent you teach online or your institution's processes are digitised; when we consider learners' learning journeys over an academic term, these journeys share the following characteristics, illustrated in Figure "Overview of learners' learning journey: an institutional perspective":

- pre-registration enquiries from learners registering for the first time (or learners re-registering)
- (for some) application processes for admission, whereby learners are screened according to set criteria before they are allowed to register
- admission to the institution, followed by registration for a particular qualification or programme and possibly a selection of courses
- access to online course sites if the institution uses a learning management system, and/or learners attending face-to-face classes for the first time
- various points of enquiry, contact and engagement as learners progress — between peers, between learners and administration, and between learners and teachers
- formative assessment opportunities
- summative assessment



Overview of learners' learning journey: an institutional perspective

Figure "Overview of learners' learning journey: an institutional perspective" provides a generalised view of how learners enrol at an educational institution, whether at the primary, secondary or post-secondary level. Although it is difficult to separate the different processes in a learners' learning journey each academic year, the figure provides a lens on three different, but most probably overlapping, phases or stages: the pre-enrolment stage, admission, and the start of the learning journey itself. The third stage is a course view and provides an overview of what happens in every course a learner takes. While stage three may have different elements in your context, this stage of the learning journey will be shaped by the number of separate courses, the delivery mode (e.g., face-to-face, online, blended, laboratory work, etc.), and engagement with institutional staff not necessarily directly linked to teaching, such as study advisors or counsellors, the finance department and/or the library.

Following from this overview is a consideration of what types of learner data are shared and collected at every stage of the learning journey. Later in this course, we specifically address different types of learner data, but for now, we want to focus on what data are collected at an *institutional level*.

Reflection action: considering learning journey data

At this stage, we want to invite you to consider your own institutional process and think about what learner data are collected at each stage of the learning journey, as illustrated in Figure "Overview of learners' learning journey: an institutional perspective". You can use below given table to record what data you know or think your institution collects from learners throughout the three different learning-journey stages.

	Pre-enrolment, first contact and re-enrolment phase	Admission and start of the learning journey	Learners' learning journeys: a course view
What learner data are collected, required and/or provided in each of these three different phases?			

Mapping the institutional collection of learner data

We suspect that though there may be some differences between contexts or institutions, if you have an opportunity to compare your table with those of other teachers, you may notice a number of similarities. What makes this activity interesting is considering whether the data that are being collected are stored as digital information, or on paper in some filing system, or whether some data (e.g., registration data) are digital, but your observations in the classroom or online spaces are located in paper-based reports. Formative assessments may be done on paper, with only the marks being recorded digitally (or not), and the assessments may not be digitised and stored on a central server. Does it matter? Or rather, *why* does it matter?

From non-digital to digital data

Let us first acknowledge that the format and scope of collecting, analysing, storing and using learner data are dependent on the context, the extent to which the institution's systems are digitised, and to what extent teachers' own processes of observation as well as formative and summative assessments are digitised.

Though there are numerous challenges and risks involved in collecting, analysing, storing and using digital learner data, or digitising learner data (from paper to digital), the advantages of having data in a digital format tend to outweigh the challenges and risks. Below given table provides an overview of some of the challenges and affordances associated with digital, digitised and non-digital data.

	Challenges and risks	**Affordances**
Non-digital data (e.g., paper-based attendance registers, observation notes, handwritten assignments and feedback)	Can be destroyed and/or lost. Difficult to share and/or expand. Difficult or impossible to combine with other digital or non-digital data. Analysing the data is more-time consuming than with digital data. Difficult to combine with other sources of non-digital information.	Often less cumbersome to write down information or observations than to record the data digitally. Existing non-digital systems are often legacy systems with vested interests in resisting change.

Digitising data from non-digital formats to digital formats	The data may be wrongly digitised due to human error. The data may be intentionally wrongly recorded. Non-digital data may be interpreted in the process of digitisation. Digitising non-digital data can be extremely time-consuming and labour-intensive.	Digitising non-digital data makes the data more usable, accessible, shareable and expandable.
Digital data	Digital data have a history that is often not acknowledged or taken into account. Digital data may be held in an obsolete format. Digital data can be stored on local or cloud servers – each with its own liabilities and risks. Digital data may become corrupted.	Digital data stored in the same formats are easy to add to, merge with other data sources and share with a range of stakeholders.

Challenges and affordances of various formats of learner data

Summary and Conclusion

This unit provided a broad, historical overview of how institutions and teachers have collected and used learners' learning data for a variety of purposes, most importantly to evaluate whether learners are progressing (e.g., formative assessment) or have achieved the envisaged learning outcomes (e.g., summative assessment). The reality is that although many institutions have moved online with regard to admissions, registration and teaching, several still use a variety of non-digital and digital formats to capture essential learner data. Even where institutions' processes (administrative and teaching) are totally online and available in digital formats, there are instances where data are recorded in non-digital formats. It is therefore important to think in terms of the whole range, from non-digital to digital data, and their associated challenges, risks and affordances.

We hope that you enjoyed this unit. In Unit 2, we explore the notions of evidence and data.

Check your progress

Congratulations on completing the first unit. Well done! Some of you may have found this overview almost too basic, while others may have had difficulty thinking about learner data from an institutional perspective. Whatever is the case, the following ten questions require you to choose only between “true” and “false,” and in some cases, it may be difficult to choose. Some of these questions go beyond the content covered so far and require you to think differently or more deeply than just finding the answer in the content above. The answers to these questions, along with further comments, are provided at the end of this unit.

a. Learning analytics is unconnected with how educational institutions have always collected and used learner data.

   i. True
   ii. False

b. Digitising non-digital data is an easy and fast process.

   i. True
   ii. False

c. There are no risks in storing digital data.

   i. True
   ii. False

d. Digital data are more trustworthy than non-digital data.

   i. True
   ii. False

e. Learner data collected for admissions and registrations have no impact on teachers' pedagogical choices and assessment strategies.

   i. True
   ii. False

f. Having access to admission, registration and demographic learner data may assist teachers with teaching more effectively, appropriately and ethically.

   i. True
   ii. False

g. Digitising non-digital data is error free.

   i. True
   ii. False

h. If teachers have access to admission, registration and demographic learner data, they can misuse the data.

   i. True
   ii. False

i. Data collected by teachers in their classes (e.g., attendance patterns, login details, teacher observations, and formative and summative assessment results) should not become part of learners' historical data files because of the dangers of misuse and labelling.

  i. True
  ii. False

j. If teachers have access to more learner data, they can teach and support learners more effectively.

  i. True
  ii. False

k. The last envisaged outcome for this unit involved you thinking about a particular course you teach. Consider the data you use to make informed decisions on issues of pedagogical design, assessment and learner support. What data do you have access to that you can use more optimally to inform your teaching?

UNIT 2

Understanding Evidence and Data

Introduction

In Unit 1, we explored an institutional perspective on learner data, from pre-admission inquiries for first-time registrations or re-enrolments, to the registration process and the start of the learning journey. We also considered the formats that learner data are collected in — digital or non-digital — and their implications for making sense of learners' learning. Having an institutional perspective on the collection, storage, analysis and use of learner data is a much bigger topic than has been addressed in Unit 1, and as we proceed, you will find different aspects of this institutional view on learner data in all of the units.

In this unit, we situate the collection, analysis and use of learner data against the backdrop of our assumptions and beliefs about evidence and data. Evidence and data have become increasingly important in developing policy, in helping us respond to pandemics or other crises, and in steering strategic and operational planning for educational contexts. So we cannot make sense of learning analytics if we do not also understand the paradigm and practices of evidence-based management, and the limitations, use and role of evidence. We also need to acknowledge some core assumptions about data — e.g., data as "raw," "complete" and "objective."

Learning outcomes

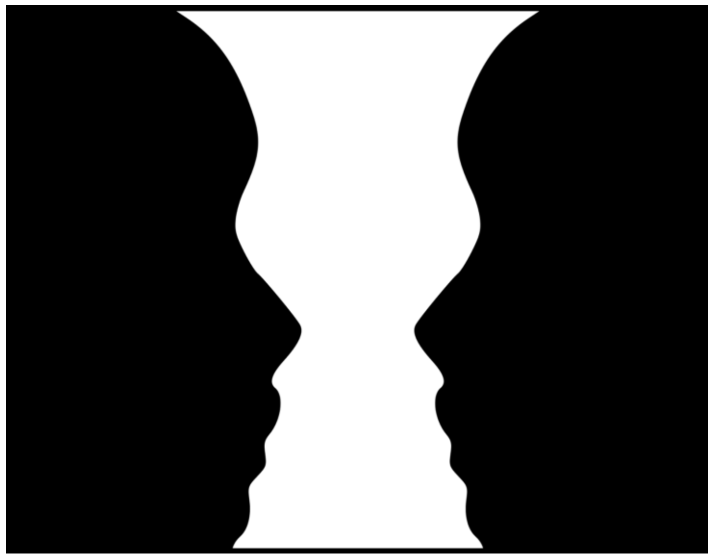
After you have worked through this unit, you should be able to:

- describe the basics of scientific evidence and the claims made in the name of research, as well as the criteria for valid claims
- critically engage with some of the general claims made about data and evidence, and understand the (mis)use of evidence in the public domain
- differentiate between causality and correlation
- define evidence-based management (EBM) in education and be able to name the limitations and dangers of EBM in education (e.g., randomised control trials)

A matter of perspective

The image in Figure "A matter of perspective" is a very popular one for explaining the power of perspective. Do you see a wine cup or two faces?

Imagine seeing two faces but the person next to you seeing the wine cup. How would you convince them that there is no wine cup? Would it be possible for them to convince you that it is, in fact, a wine cup?



A matter of perspective

By analogy, it is possible that you and your colleagues can look at the same piece of evidence but have two different opinions on what it means. While in this example, the difference in perspectives is relatively trivial, conflicting views on the same phenomenon can be much more serious. For example, think about public debates over whether immunisation causes autism, whether climate change is real, whether we can trace the history of humankind to evolution, or whether there is a life after death. In many of the debates surrounding these issues, each group will have their own set of evidence, shored up by science or their subscription to a set of beliefs (religious or otherwise). Making matters worse is the phenomenon of "fake news," which often contains enough fragments of truth to lure the reader in, and is used by many to label a particular claim or belief as false, notwithstanding the validity of the evidence.

Welcome to the world of data and evidence — a world that may seem like a minefield at first. But don’t give up! In the rest of the unit, we will examine a number of guiding principles to help make sense of various claims and counterclaims.

The notion of “raw” data

Thinking about the notion of “raw” data suggests that if data can be “raw,” then it can also be “cooked,” “cleaned,” “processed,” “sanitised,” etc. An example might be if you record an interview with a learner or a teacher, and on playing it back, you want to eliminate the background noise, the pauses, the interruptions and the times when your interviewee hesitated. You may also want to exclude from your report, views or statements that do not support your views or your claims. By the time you present your findings, they are based not only on what you found, but also on what you excluded.

The original recording of the interview is often referred to as “raw” data — meaning data that were found, data that have not been processed, cleaned and formatted. But was that data really “raw”? Imagine that you invited the colleague for an interview, she asked you what the interview was about, and you provided her with a broad outline of what you hoped to achieve. You may have selected her based on a number of criteria (e.g., you were looking for opinions of black females older than 40 who are employed in executive positions), but you also had in mind a number of set questions (whether open-ended or semi-structured) that would guide the conversation.

You may have allocated an hour for the interview and arranged for it to take place in a coffee shop. And then you conduct the interview, which you record on a portable voice-recorder. Afterward, you save it as an audio file. When you listen to the audio file, you use software to erase the background noise of the coffee shop, you erase the small talk about her rush to get to the interview in time, and you erase your apologies for being late for the interview because you were stuck in traffic. In transcribing the interview, you do not include her hesitations, her facial expressions or her body language.

Can we really speak about the initial data as “raw”? A number of decisions preceding the recording of the interview shaped the recorded data in ways that suggests they are not raw. Of all the possible candidates, you selected your interviewees based on some criteria, including availability. And what if, although you did not mention it explicitly, you did not consider the available candidates who were female, black, over 40 and in executive positions but were also living with physical disabilities, or were in established heteronormative relationships?

In such a scenario, can we really consider the recorded interview as “raw”? We don’t think so. We hold the position that the decisions that frame the research focus or question, the research methodology, the sample, and the exclusions and inclusions go on to shape the data in a number of ways, and that we cannot and should not talk about data as “raw.” If we also take into account the processing that followed the original recording — the reduction of background noise, the deletion of introductions and small talk, the deletion of hesitations, and the inability of the audio recording to capture facial expressions and body language, we are even more convinced that the final data set was anything but “raw.”

Consider the following example: You are curious about factors that influence learners' chances of success. You decide that you will analyse learner participation data in an online discussion forum where participating is not compulsory. Secondly, you decide to look at participation in a selection of topics rather than all of the topics. To help you make sense of responses, you also define "participation" as posts containing at least ten sentences. Once you have selected the content based on the above criteria, you proceed with the analysis of the "raw" data. But again, you cannot speak of this data as genuinely "raw," given the processes preceding the collection and how these influenced what data were collected and what were excluded.

Making sense of knowledge claims: trustworthiness in research

There is a general belief that “numbers don’t lie” and that one can trust hard statistical evidence about a particular phenomenon more than, for example, perceptions or impressions. Let us consider the following: You are presented with two sets of evidence on the number of learners who have passed an online course. One set provides a table of the number of learners who attempted the examination and achieved a minimum of 50% (the qualifying pass mark). This table indicates that 65% of learners who attempted the examination passed with at least 50%. The other set of evidence contains interviews with one of the teachers and one of the markers. The teacher indicates that she has not seen the final data yet, but from her experience and from how learners were doing in the formative assessments, she thinks about 80% of learners will pass the examination. The marker indicates that he was allocated half of the submitted papers to examine, and from assessing that half, he estimates that the pass rate should be about 75%.

Which of the two sets of evidence would you trust? This is a fairly easy example — the one set of evidence comprises calculations of the number of learners who wrote the examination, their marks and a 50% cut-off point. The other set contains impressions and guesses. So when someone asks you how many learners passed the examination, you will most probably rely on the first data set.

But before you think this is simple, consider the possibility that the head of the department or your manager calls you in and asks you the same question, and when you provide her with the answer, she responds by saying, “I want to know what percentage of learners who enrolled for the course passed the exam.” Now, this is a different question. The first data set only considered learners who attempted the examination. It is well known that many learners in online courses drop out before the examination or do not attempt to write the examination. So if 100 learners registered for the course, but only 80 attempted the examination, then the total pass rate for the course looks different from the pass rate of those who sat for the examination. Though 65% of the 80 learners who sat for the examination passed, if one takes the initial number of learners into account (100), the pass rate is even lower: 52%.

The point we are trying to illustrate is that whilst numbers or statistics are regarded as more truthful, one should take care to look at how the inquiry was phrased!

General principles of scientific research are that the research should be valid, reliable and generalisable. Note that these three criteria do not apply in the same way to research based on qualitative data, such as interviews, open-ended surveys, and/or observations. However, let us start with the three criteria that apply to quantitative research (research using numbers, like the example above).

- **Validity** refers to whether the concept being measured is measured accurately. For example, if your manager asked you for the pass rate of the whole enrolment, but you provided her with the pass rate of those who sat for the exam, this would not be valid. In another example, if you want to measure how many people suffer from depression, but your questions actually measure anxiety, your research will not be valid.
- **Reliability** refers to the consistency with which another researcher will find the same results. For example, if another researcher reports on how many learners from the original enrolment passed the exam, she should find the same answer if she uses the same method of calculation.
- **Generalisability** refers to whether we can apply the research results of a particular calculation to other contexts, using the same method and finding the same results. For example, consider asking about the gender, race or age of those learners who sat the exam and passed, and then claiming that in all online examinations in your subject, a particular gender, race or age would always perform better than those of a different gender, race or age. Based on the example we have used so far, you would not be able to make that claim, so your findings would not be generalizable. However, if your calculations were based on a national online exam and you used data from the last ten years, the potential generalizability of the findings would be much better.

With regard to qualitative research (e.g., interviews, focus groups, observations), different criteria allow us to judge whether the findings are *trustworthy*. You will immediately notice that we use different terminology for qualitative research. As we illustrated, the use of numbers or statistics does not necessarily mean the findings are valid, reliable and generalisable. Conversely, findings that are *not* numerical or do not contain statistics are not necessarily untrustworthy. Below are four criteria to consider when evaluating the trustworthiness of qualitative findings.

- **Credibility** refers to activities that confirm your analysis is credible if another independent researcher confirms the analysis and findings based on a subset of your data. The data not analysed then serve as checks and balances to see whether the findings are still valid.
- **Transferability** refers to whether the findings can be used in another context. This is similar to the criterion of "generalisability" in quantitative research. In qualitative research, it is not the researcher who claims that her findings are transferable; instead, other researchers engaging with her findings and her detailed, "thick" analysis decide whether the findings also apply in their respective contexts.
- **Dependability** refers to how reliable the analysis is. Dependability can be ensured by using multiple methods, sources, theories and fellow researchers to investigate and analyse a phenomenon; this is also referred to as "triangulation." Researchers can also ensure dependability by being transparent about their processes and documenting these so they can be audited if necessary.
- **Confirmability** links to dependability and can be enhanced by the researcher making field notes, recording observations and keeping a journal.

We realise that the above may be too much (or not enough) information if you have never thought of data and research in a more systematic way. If you are curious or want to know more, explore research methods **here** and **here**.

Understanding causality and correlation

One of the major dangers in looking at student engagement patterns, for example, and then considering whether learners who participate more in online discussions also have a higher pass rate is that we will confuse causation and correlation.[2] *Causation* refers to a cause-and-effect relationship, such as a stone thrown in a pool of water causing the water to ripple outward. There is a direct link between throwing the stone into the water and the stone's impact on the water causing ripples. It is important to note that causation implies there is no alternative explanation for why there are ripples in the water; they arise specifically because someone has thrown a stone into the water. *Correlation* is when we see the ripples in the water, but we have not seen anyone throw a stone in the water; there are kids playing close by, but there are also fish coming to the surface, frogs who may have jumped into the water, and insects who may have landed on it. You cannot be sure which one of those caused the ripples in the water, but you can say there is a possible correlation between seeing ripples in the water and having one or more of those other circumstances occurring at the same time.

Let us now consider causation and correlation in teachers' attempts to improve learners' chances of success. One teacher may believe it is the extra classes he provides that are making the difference. Another teacher believes learners' marks have improved since she started to use a particular technology in the classroom. Teachers often believe that whatever strategy they employed has caused the change. They frequently report on the effectiveness of particular strategies by indicating learners' performance before the intervention, describing the intervention, and then presenting measurements of the learners' performance after the intervention. If there is an improvement in learners' performance, these teachers claim their strategy caused the improvement.

[2]See this explanation of the difference between causation and correlation.

There is an established research methodology first used in scientific (specifically medical) research that has gained some popularity in education: randomised controlled trials. In medicine, researchers investigate the impact of one factor or variable on a selected organism or portion of an organism to study its effects. Important in such laboratory settings is that all other variables that could also influence the impact of the factor are excluded. Because the only thing that changed was the one factor or variable, researchers are then able to explain the changes that occurred as being direct results of the introduction of the one variable. They can then say, without a doubt, that the introduction of the one variable caused these changes. Even more important is that researchers can then claim that under exactly the same conditions, the introduction of the factor or variable will have exactly the same outcome. This is also referred to as the reproducibility of the results.

Think of a classroom situation where you randomly divide the class into two sections (meaning everyone has an equal chance of being in either section). You expose one group to an intervention, and the other group (known as the control group) continues as normal. Afterwards, you have all the learners do an activity, and you measure whether the group that was exposed to the intervention performs differently from the control group. Analysing the results, you then claim that the intervention caused one group to do better than the other group. Such an experiment would be a randomised controlled trial (RCT).

While such an experiment seems relatively easy (and harmless), a number of challenges may have an impact on the claims made during an RCT. For example, if the experiment of introducing a selected group of students to a particular technology takes place over a long period of time, it becomes impossible to exclude other factors or variables.

What if during these interventions, the school also starts offering free meals, renovates classrooms, or begins providing disadvantaged learners and their families with psychological and social support? While the teacher's intervention may have had an impact, we cannot say for certain that the intervention caused the improvement in learners' performance. There may be a correlation between this intervention and the improvement in results, but the intervention was not the only element in the school environment that changed during this period. The improved results could therefore be due to a combination of factors.

Education as an open and recursive system

The preceding section is a wonderful introduction to considering education as an open and recursive system.[3] What do we mean by "open and recursive"?

There are many aspects within the control of teachers in a teaching and learning environment, e.g., pedagogical choices, assessment strategies, quality of assessment feedback to learners, and so forth. There are also many aspects in that environment controlled by administration and management, such as budget and resource allocations, planning, and so forth. It is fair to say, though, that there are many things outside the control of those directly linked to the institution — for example, economic conditions, crime in the areas surrounding schools, the employment status of parents and carers, and other factors. Education should therefore be considered an "open" system, as it is impossible to isolate a classroom from the rest of learners' lives — their communities, the circumstances affecting parents, and so on. The term "recursive" refers to the ways in which those patterns of influence tend to repeat themselves.

[3] Biesta, G. J. (2010). Why "what works" still won't work: From evidence-based education to value-based education. *Studies in philosophy and education*, 29(5), 491–503.

As we will explore in the next section, we often think of classrooms (whether physical classrooms, online, or hybrid) as laboratories where we can isolate learners from outside influences, change a single aspect of their learning, and observe the impact of that change. This belief that one can isolate learners and change one variable while not changing anything in another group of learners underlies the notion and practice of randomised control trials.

Though many things are within the control of teachers and learners, there is much that is not within their control. While it is understandable for teachers to claim that their interventions have made a difference, we should be careful not to confuse correlation with causation. If you are curious about how the confusion between correlation and causation can play out, here are some examples: there is a correlation between the consumption of margarine and divorce rates in Maine (France);[4] there is also a correlation between autism and living close to a freeway.[5] In the latter case, a newspaper headline presented this correlation as causation.[6] In a 2015 article, the Harvard Business Review[7] discusses the dangers of thinking that because there is a link between two phenomena, this means the one causes the other.

[4] https://www.datasciencecentral.com/profiles/blogs/spurious-correlations-15-examples

[5] https://www.latimes.com/archives/la-xpm-2010-dec-16-la-he-autism-20101217-story.html

[6] https://www.statisticshowto.com/spurious-correlation/

[7] Harvard Business Review. (2015, June). Beware of spurious correlations. Retrieved from https://hbr.org/2015/06/beware-spurious-correlations

Evidence-based management and teaching

This may not be news to you, but the emphasis on using evidence to inform educational strategy and implementation is a global one. Governments, regulators, and management structures of educational institutions all look for “what works.” While it is understandable that we should be concerned whether the resources allocated to interventions, new laboratories and new staff make a difference and “work,” a number of authors, such as Biesta, [8, 9] argue that “what works” is not necessarily the right criterion. Rather, Biesta suggests that we should also consider whether the intervention or action is “appropriate.” An example of an action that “works” but that may not be appropriate would be to offer extra classes on the school premises to help increase the success rate, but only for those kids who can pay. While we all would agree that extra tuition would most probably “work” and raise learners’ performance, the prerequisite that learners must pay to attend extra classes, offered on the school premises by the very same teachers responsible for the classes during school time, is definitely not appropriate.

[8] Biesta, G. (2007). Why “what works” won’t work: Evidence-based practice and the democratic deficit in educational research. Educational theory, 57(1), 1–22.

[9] Biesta, G. J. (2010). Why “what works” still won’t work: From evidence-based education to value-based education. Studies in philosophy and education, 29(5), 491–503.

Summary and Conclusion

We don't doubt for one second that evidence plays a seminal role in making better, more effective and more appropriate decisions. But as we have pointed out in this unit, we need to critically engage with that evidence, whether it is based on quantitative or qualitative research.

Just because something is presented in numerical form or as a compelling narrative does not mean that the evidence is trustworthy.

Learning analytics is becoming increasingly central to looking for evidence around why some learners are more successful than others, or determining whether our interventions as teachers make a difference. This unit has laid the foundation to celebrate the huge potential of learning analytics to help us teach better, and help our learners to learn better, whilst recognising that we should also critically engage with claims made based on those analytics.

Check your understanding

At the outset of this unit, we invite you to reflect on the following questions. It's fine if you're unsure or don't know the answer. As you read the unit, you may recognise some of the issues raised in these questions. We also provide pointers and further explanations at the unit's end.

a. Think back to the previous unit and how institutions collect learner data. We emphasised how formative and summative assessments are ways to ascertain whether learners are making progress and have achieved the envisaged learning outcomes. Consider the following statement: The marks learners receive in a summative assessment (e.g., a written examination) are a true reflection of whether learners have achieved the desired outcomes.

   i. True
   ii. False

b. There are many aspects of learners' learning that we cannot capture, which means our "evidence" of their learning is always partial and tentative.

   i. True
   ii. False

c. We can consider data about learners' performance to be neutral and a true and complete version of their performance.

   i. True
   ii. False

d. Evidence suggests that learners who study in their home language have a greater probability of passing their course than learners who study in a second or third language. The relationship between studying in your home language and passing a course is an example of causation, not correlation.

   i. True
   ii. False

e. In testing the effectiveness of an intervention to improve learner success, we can divide a class into two groups: one that will be exposed to the intervention, and one that will not (the control group). If the group receiving the intervention does better than the control group, this is proof that the intervention will work for all learners in that disciplinary and institutional context.

   i. True
   ii. False

f. Having interviews with learners and their parents about the effectiveness of a particular technology in teaching and learning is an example of quantitative research.

   i. True
   ii. False

g. Instead of conducting interviews with learners and their parents, you circulate a survey on the impact of technology on teaching and learning. Your questions ask respondents to use a number (1, 2, 3, 4 or 5) to indicate how important they think a particular aspect is, or to what extent they agree with a statement, where 1 indicates they do not agree at all, 2 they disagree somewhat, 3 they do not really have an opinion, 4 they agree somewhat and 5 they agree completely). You will calculate the responses and will be able to report on how many of your respondents indicated that they agreed, agreed somewhat, did not have an opinion, disagreed somewhat or disagreed completely. This is an example of quantitative research.

   i. True
   ii. False

h. The notion of “transferability” in qualitative research means your findings from the interview with parents and learners can be applied to another context.

   i. True
   ii. False

i. How will you determine whether people find your quantitative analysis of the survey on the impact of technology on teaching and learning reliable? (Choose one.).

   i. if my headmaster or head of department confirms that they agree with the findings
   ii. when the findings are used in another context
   iii. by ensuring that if another researcher uses exactly the same questionnaire and the same methods in another context, they will find the same results

j. A reporter visits your institution to investigate the work conditions of teachers, and she only speaks to the principal or head of the institution. She then reports her findings and claims that teachers are very satisfied with the working conditions, not mentioning she only interviewed the principal or head of the institution. What is wrong with such a report?

   i. The reporter should also have spoken to the kitchen and maintenance staff.
   ii. Choosing an interview as a data collection tool was not appropriate; she should have sent a quantitative survey to the headmaster to get her responses in writing.
   iii. She should have at least indicated in her report that she only spoke to the principal, or she should also have interviewed teachers.

If you found these questions difficult, welcome on board! When we get to the end of this unit and revisit the questions, you may be in a better position to consider your initial responses and, possibly, change your answers.

UNIT 3

Understanding Academic, Learning and Teacher/Teaching Analytics

Introduction

Welcome to Unit 3! In Unit 1, we discovered the world of learners' data from the perspective of their enrolment at an institution — and how different data are collected at various stages of their learning journey. We started to explore how teachers can use available data to inform their pedagogies, assessment strategies and learner support. Using learner data is, however, fraught with potential risks and misinterpretation, as we discovered in Unit 2!

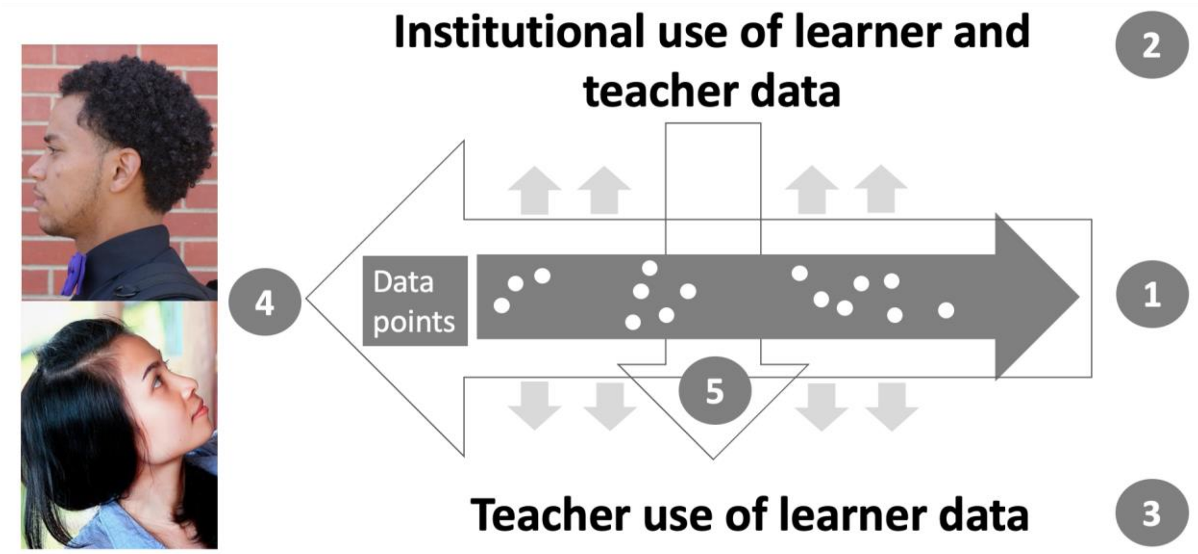
Unit 1 also made us aware that while teachers collect and use learner data for a variety of purposes, so does the institution. Although teachers have an interest in measuring learners' progress and identifying those who may need support or extra stimulation, institutions collect and use learner data to plan, allocate resources and report to a range of stakeholders on a variety of issues, such as the quality of the offerings, trends in enrolment, student success, and so forth. This, in very simple terms, is the main difference between *learning analytics* and *academic* analytics. In addition, there is an emerging variation known as *teacher* or *teaching* analytics.

Learning outcomes

After you have worked through this unit, you should be able to:

- distinguish between learning analytics, academic analytics and teaching/teacher analytics
- define learning analytics and formulate a personal definition of learning analytics for your local context
- provide examples of the different types of analytics in your own course/institution

Using learning journey data: an introduction to the players



Use of learner data [10]

So far, we have focused on two main players in the use of learner data: the institution itself and teachers. We have seen how institutions collect and use learner data, and we now have a sense of how teachers collect and use learner and learning data. The illustration in Figure "Use of learner data" creates a visual representation of different aspects of the use of data in an educational context.

On the left, we have two students who have enrolled at the institution and started their learning journeys through the different phases discussed in Unit 1. First is admissions and (re)enrolment, registration and study.

[10] The two pictures of students were obtained from Pixabay, where they were published free for commercial use, no attribution required. Retrieved from https://pixabay.com/photos/side-profile-black-male-student-1440176/ and https://pixabay.com/en/girl-library-education-student-1721436/.

This journey is indicated with the arrow from left to right (1). As students progress through the different stages, they leave digital and non-digital traces and evidence of their learning journey (indicated by the dots in the arrow from left to right). These data points can then be collected by the institution (2) and teachers (3) for different uses (see section 3.3). The personal and behavioural data collected by the institution and teachers are analysed, and insights are shared with learners to help them make more informed and appropriate decisions (indicated by the arrow from right to left (4). We will discuss this aspect in section 3.3.

There is, however, another aspect that we might consider. Institutions have always monitored and evaluated the performance of teachers, and as education is increasingly moving online, institutions are beginning to use teaching and teacher data (the arrow pointing down, 5) to monitor and evaluate teacher performance.

Making sense of learning, academic and teacher/teaching analytics

In this section, we discuss further the three broad categories of analytics: learning, academic and teacher/teaching analytics. These three categories should not be confused with types of data (see Units 4 and 6 for discussions on sources and types of data). We also acknowledge that there are (increasingly) other forms of analytics for various aspects of teaching and learning that claim unique domains separate from these three forms — e.g., class and course analytics.

Central to academic, learning and teaching/teacher analytics are *data*. These data originate as learners are admitted to the institution, register, and start their learning journey, having chosen a particular programme and/or courses. Learners encounter the teachers allocated to these courses, who have the responsibility to teach, support and assess learners as they achieve the course learning objectives. The institution has, as its mandate and fiduciary duty, the responsibility to oversee and enable learners' learning journeys, and also to hold teachers accountable for the execution of their duties.

Below given table provides an overview of the three types of analytics: the data used, who uses them, and for what purposes.

	Academic analytics	**Learning analytics**	**Teacher/teaching analytics**
What data are collected?	Demographic data Individual learner summative assessment data Class and/or cohort data	Demographic data Individual behavioural learning data (e.g., attendance, engagement, observations) Formative assessment data Summative assessment data	Observation data through class visits (face-to-face), or online engagement data Data provided by teachers on their performance Peer evaluation

Who uses the data?	Management of schools/ institutions Provincial and national governments Quality assurance bodies Funding agencies Researchers	Teachers Learners Learner support teams Course teams	Heads of departments or sections Institutional managers
Why are the data collected?	Strategic and operational planning Resource allocation Quality assurance Reporting purposes Research	Check progress Predict outcomes Provide personalised support and guidance Gain insight into effective learning design	Performance management Quality assurance

Overview of academic, learning and teaching/teacher analytics

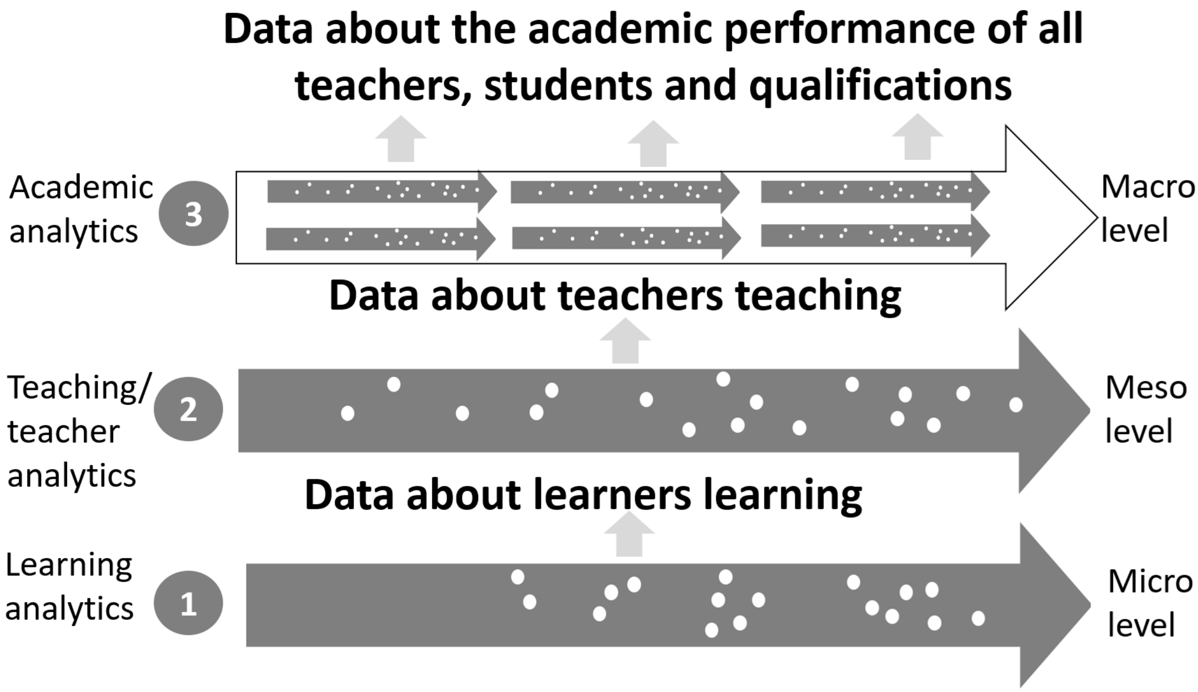
Academic analytics: Institutions do not tend to focus on the details of what is happening in classes, or what is happening at an individual level; they are largely concerned with the outcomes of the learning journey, that is, whether students pass, fail or drop out. Institutions might be expected to focus their data collection on learners' final marks (allowing them to progress to the next level or to graduate), and on how a class or group of learners enrolling for the same course have performed. Institutions use these data, among other things, for reporting to provincial or national governments to justify funding, reporting on the quality of their programmes, marketing, as well as strategic and operational planning. National governments are also interested in researching trends in school-leaving examinations, compiling databases of data from different schools for national overviews, and assisting schools in measuring expected learner progress. For example, in the UK, the Fischer Family Trust [11] analyses cohort-wide learner outcomes annually, providing insights to schools for further focus and improvement.

[11] https://fft.org.uk/

Learning analytics: Learning analytics differs from academic analytics in being focused, unsurprisingly, on learners, particularly at the course level (e.g., Economics 101), and on the *individuals* in the course — what they need, how they are progressing and what can be done to better support them. While learning analytics may use some of the information collected by the institution (e.g., demographic data), learning analytics pulls in additional types and sources of data, such as class observations (face-to-face and/or online), formative assessments and summative assessments. The aim of learning analytics is not to report what happens in Economics 101 to national governments or regulatory bodies, but to teach more effectively and to support students more appropriately. It is also hoped that sharing the findings of learning analytics with learners may help them to make more informed choices.

Teacher/teaching analytics: Teacher or teaching analytics has links with learning analytics in the sense that both aim to ensure teaching and learning are as effective and as appropriate as possible. But whereas learning analytics has learners as a primary data source, teacher/teaching analytics focuses on what teachers do in the classroom, how they perform, etc. In some institutions, teacher/teaching analytics may be applied in service of the contractual agreement between the institution and teachers to ensure that teachers perform to expected standards.

We might also think of the three types of analytics as occurring on three different levels, as represented in Figure "An overview of academic, learning and teacher/teaching analytics". Learning analytics (arrow 1) functions at a micro-level, with each course having its own set of analytics to inform teaching and provide feedback to teachers and learners. The middle arrow (2) refers to teacher/teaching analytics, which provide the institution with feedback on how teachers are doing in terms of agreed standards. The top arrow (3) refers to academic analytics; you'll notice immediately that this is actually a collection of arrows, each referring to the different courses and qualifications offered at the institution.

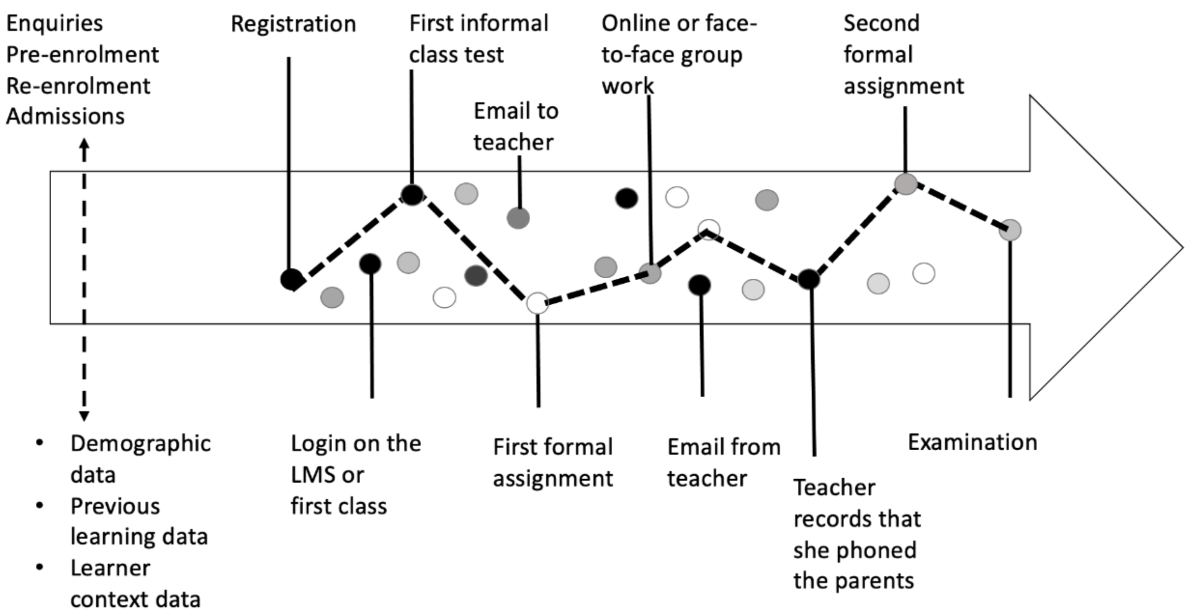


An overview of academic, learning and teacher/teaching analytics

Defining learning analytics

While learning analytics developed from a range of disciplines, such as education, psychology, computer science, mathematics, and data science, its central concern has always been about learning. The first Learning Analytics and Knowledge Conference (LAK’11) was held in Canada in 2011 and defined learning analytics as follows: “the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimising learning and the environments in which it occurs.” [12] Consider for a moment the illustration in figure "Visual representation of points of learning analytics" of a learning journey.

[12] https://tekri.athabascau.ca/analytics/



Visual representation of points of learning analytics

That original definition of learning analytics refers to four actions in relation to data: measurement, collection, analysis and reporting. Look at figure "Visual representation of points of learning analytics", and consider the kinds of data collected from the first stage (as discussed in Unit 1) through to the examination. Data are collected at the pre-enrolment and registration stages — for example, previous learning records and demographic data regarding the learner's context (e.g., home address, race, age, gender, etc.) — but also data that emerge as the learning starts, such as a first class or login on the institutional learning management system (LMS), forum posts, email communication to and from the teacher, informal class activities and observations, formative assessment (compulsory assignments), a note that the teacher phoned the parents, and marks from any summative assessment (e.g., examination).

The definition is clear that the data are collected "for purposes of understanding and optimising learning and the environments in which it occurs." The data should therefore help us first to *understand* learning (more about that in Unit 5), and then to *use* this understanding to optimise learning and the environments in which the learning occurs.

Reflection action

In this activity, there are no right or wrong answers. We would like you to engage with the questions and formulate a position or a response.

a. Think about your own context where you teach (if you teach), and look again at Figure "Visual representation of points of learning analytics". You may be in a context that is mostly face-to-face and may record these different interactions in non-digital formats, or you may be teaching a totally online course where you have access to a range of digital data (remember the different options in Unit 2?). Depending on your context, how would you define learning analytics?

b. This unit has focused on three types of analytics: academic, learning and teacher/teaching. Refresh your memory by looking again at Table "Overview of academic, learning and teaching/teacher analytics". Thinking of your own institution, see whether you can discover information about these three types of analytics. For each category, find out what data are collected, who uses the data and why the data are collected (i.e., for what purpose).

	Academic analytics	Learning analytics	Teacher/teaching analytics
What data are collected?			
Who uses the data?			
Why are the data collected?			

Summary and Conclusion

In this unit, we explored three distinct types of analytics: academic, learning and teacher/teaching. Table "Overview of academic, learning and teaching/teacher analytics" provided a brief overview of these three types and illustrated how they each collect and use different types of data, for different purposes and different audiences. Though teachers may be interested to know broader institutional statistics, they tend to have a greater interest in analytics insight that could help them to teach more effectively and to support learners more appropriately. Just as teachers may have a class focus, the institution may not necessarily be interested in how learners engage in a single class, or download resources from the LMS, and so forth (unless that class has unique characteristics that make it of particular interest). One might say that learning analytics focuses on the micro-context of learner performance, while academic analytics focuses on the macro or institutional level.

So, what about teacher/teaching analytics? It helps to think of this as being at the meso level, in between the institutional level of academic analytics and the micro level of learning analytics.

Check your progress

a. Learning analytics provides useful insight for provincial and/or national governments.

   i. True
   ii. False

b. Teacher/teaching analytics is used primarily to inform performance evaluations of teachers.

   i. True
   ii. False

c. Both academic and learning analytics make use of student demographic and context data, albeit for different purposes.

   i. True
   ii. False

d. Which of the following is not a characteristic of learning analytics? (Choose all that apply.)

   i. Aims to optimise learning
   ii. Wants to understand learning
   iii. Is primarily used by provincial and/or national governments
   iv. Is used by funding agencies

UNIT 4

Sources of Data: Mapping the Data Journey of Learners

Introduction

Many things have changed since humans hunted for their food, or navigated without maps or GPS! While the Internet has clearly become deeply embedded in how we live, and how we make decisions, connect to others, learn and teach, we know that not everyone is equally connected. In some parts of the world, Internet connectivity may be problematic, whether due to a lack of network coverage, a lack of available devices and/or hardware, the affordability of data, or broader issues in communities, such as gender inequality, or power dynamics. The World Bank[13] published a report in 2016 showing that the affordances of the Internet are not equally spread. If you are curious about how connected the world really is, and how connected your own country is, visit this SlideShare presentation from Data Reportal **here**.

In Unit 4, we aim to map learners' data in terms of their engagement with the institution — from their first point of contact or inquiry, to registration, and eventual completion. This will provide us with an overview of the data we have, where it resides, who has access to it, under what conditions and for what purposes.

Learning outcomes

After you have worked through this unit, you should be able to:

- map the learner journey in your respective course contexts
- identify what data are captured, by whom, in what form (digital or paper-based) and for what purposes, where the data are stored, who has access to them and how the data are used at each point in the learner journey
- categorise data sources as directed, automated, gifted or contractual

[13] World Bank. (2016). Digital dividends. International Bank for Reconstruction and Development & World Bank. Retrieved from https://www.worldbank.org/en/publication/wdr2016.

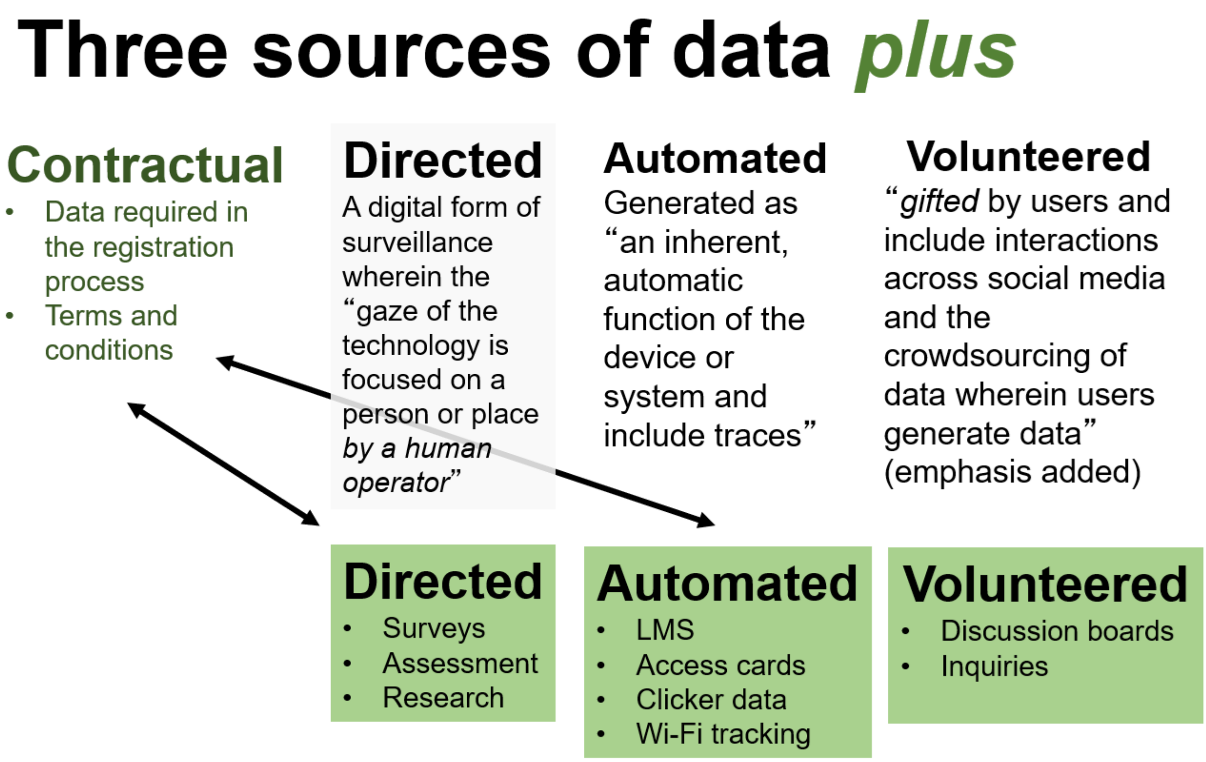
Data categories in learning analytics

Nowadays, there is a huge amount of data on the servers of our institutions. Whether big data or small, they are valuable and important. You may have come across the term digital exhaust, meaning the data gathered from Internet use or other digital activities. These comprise the fuel that enables learning analytics to function. Without data, it is impossible to understand learners' behaviour, evaluate systems, or assess courses within digital systems.

Learners are key for producing data in educational digital systems. Sources of learner data (see Figure "Four categories of data sources") can be broadly categorised as one of the following (Kitchin, 2013):

Automated data	In the 21st century, with increasing levels of automation, institutions rely on digital systems generating data from individuals' traces. Automated data are generated as an innate function of devices and systems. One example is clickstream data, which registers how individuals browse, move across digital content, etc.
Directed data	generated by systems operated by humans — (e.g., surveillance or other tracking information).
Gifted data	voluntarily provided by an individual (e.g., forum posts).

Contractual data	based on formal, legitimate and fixed agreements between parties. An example might be your agreement to accept cookies to store your information, or the compulsory data provided at registration.



Four categories of data sources

The inclusion of such a broad range of data sources has resulted in vast volumes of data being available like never before!

Data types

So what types of data can be collected or mined in learning analytics? Several studies have attempted to answer this question. Below given table provides an informative overview of data types in the field of learning analytics (Khalil & Ebner, 2015; Lin et al., 2019).

Data type	Source of data	Example
User interaction data/data traces/activity data	Automated	Mouse clicks, videos accessed, documents accessed, file downloads, votes in discussion forums, time of login
Content data	Volunteered	Video information, metadata, audio files
User profile	Contractual, Directed	Age, gender, learning interest
Contextual data	Automated, Directed	Timestamps, location
Learning records data (not applicable to all learning platforms/data sources)	Contractual, Automated	Student historical rating for constructing a recommender system
Academic data	Directed	Previous studies, grades, exams taken, certificates

Data types in learning analytics

A key characteristic to keep in mind when working with learning analytics is to always remember that it is about learning and optimising the learning experience of learners, regardless of what data types and data categories you acquire!

Educational level data

It can be remarkably difficult to compare a measure developed in one context against those developed in another. Education systems vary considerably across the world, and this can make it hard to draw meaningful comparisons. In response, UNESCO initiated a framework [14] for organising education programmes and qualifications by applying uniform and agreed definitions to facilitate comparisons of education systems across countries.

As well as this, the increasing but mixed use of online platforms leads to huge differences in data captured for learning analytics. Educational data sets can vary in size, degree, level and complexity. Currently, learning analytics has most often been applied to:

- school contexts (e.g., K-12), including early childhood education (nursery/kindergartens), primary, lower secondary and upper secondary education
- further/higher education, including colleges, undergraduate degrees, masters, and PhDs
- vocational education and training, including preparing learners for the world of work

[14] http://uis.unesco.org/en/topic/international-standard-classification-education-isced

Examples of educational-level data and learning analytics

Educational level	Digital platform	Learning analytics applied
Schools, e.g., K-12	Social network platform	Teachers may track learner progress against curriculum milestones and use test results to identify learners who may benefit from additional support. In schools adopting greater use of online platforms, staff may measure learner engagement by collecting information on number of posts, etc. to understand at a glance what is happening in classrooms across their schools. Some schools have even adopted tools for behavioural analytics, although these are not free from concern.[15]
Higher education	Learning management system	Tools can be used to flag a learner falling behind, e.g., by failing to attend classes or submit an assignment, or to predict learners potentially at risk. Interventions can take the form of calls or emails or the use of dashboards or recommender systems
Virtual learning platform	Virtual learning platform	Teachers use dashboards to inspect class progress and monitor individual learning processes

Examples of applied learning analytics for different educational levels and digital platforms

[15] See Clarity Innovations. (2015). From analytics to adaptive learning: An overview of K-12 business models and opportunities, page 24. https://www.k12blueprint.com/sites/default/files/AnalyticsWhitePaper-FINAL.pdf.

Data sources

There is a plethora of data sources for learning analytics. With the ubiquitous spread of technology across education, including growing uses of artificial intelligence and machine learning, there is a potentially vast collection of data available for analysis. Computer-based systems are a positive goldmine of learner data. For example, even simply logging timestamps of login frequencies or tracking the coordinates of mouse movements can produce a massive amount of data for an average class.

Since this chapter highlights learners at a range of different educational levels, we have tried to provide a list of the most common educational computer-based systems where data are generated. Learning analytics data may be mined from many learning systems, including those listed in below given table.

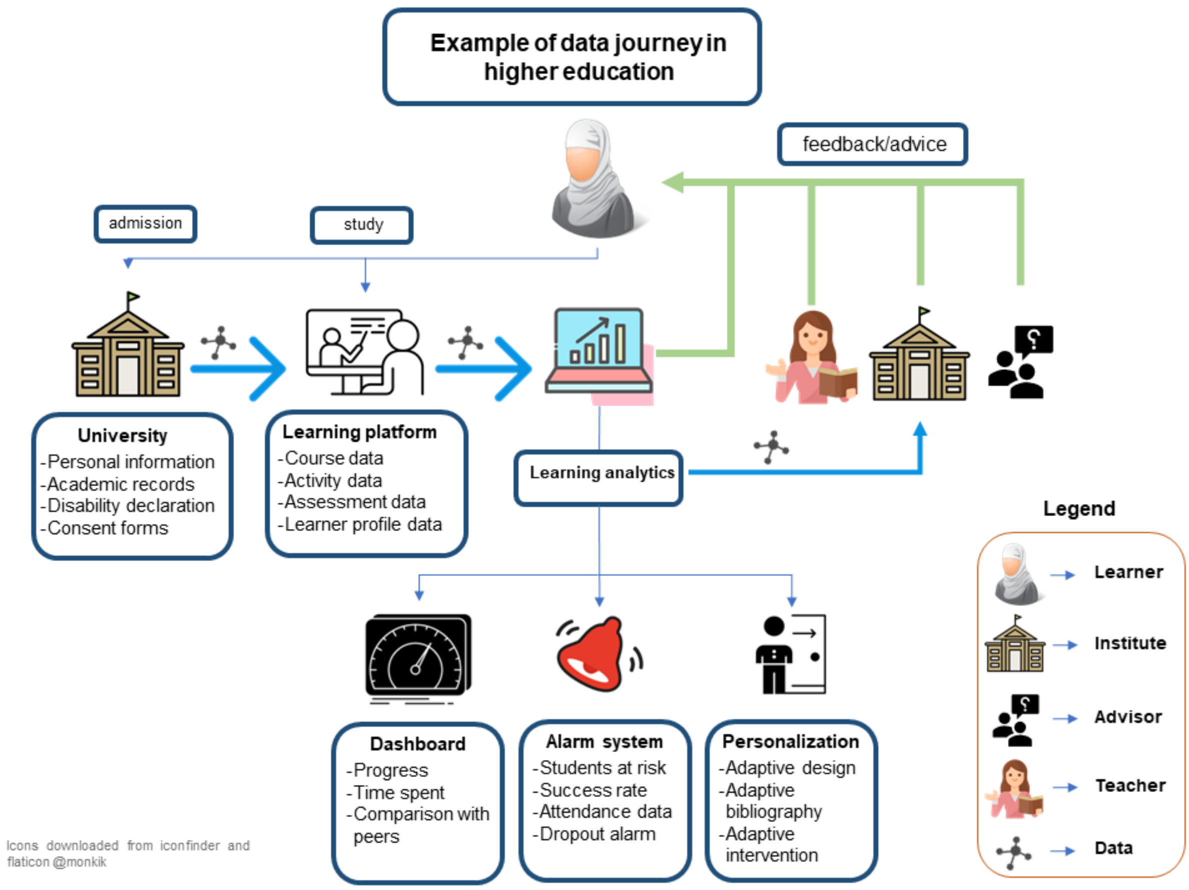
Data sources	Description
Learning management system	This is a software application for the administration, documentation, tracking, reporting, automation and delivery of educational courses or training programmes. Common software packages used in educational institutions include Canvas, Moodle, Desire2Learn and Sakai.
Massive open online courses (MOOCs) and small private online courses (SPOCs)	A MOOC is an online course aimed at unlimited participation and open access via the Web. MOOCs are often free and may be educational, vocational or for interest. A SPOC is a form of MOOC but is typically used locally and aimed at on-campus students. Examples of providers include edX, Canvas, FutureLearn, OpenHPI, Miríadax and MITx.

Intelligent tutoring system (ITS)	Provides adaptive instruction and feedback to learners through the use of learner behaviour modelling. Examples include Active Math and Cognitive Tutor.
Other types	Multimodal systems such as WEKIT, virtual reality and augmented reality systems, social media, wikis, discussion forums, different forms of interactive data generation systems.

Data sources from computer-based systems in learning analytics

Mapping learner data

Getting a better understanding of timing, places and activities in relation to learners is paramount when the objective of the technology use is to offer a better learning experience. In this part of Unit 4, we provide a short case study of a learner's data journey.



An example of data journey in learning analytics

It can be helpful to use visualisation techniques to map learner data journeys (Howard, 2014). Let us imagine this represents the journey of one learner at a university who makes good use of online data. At the point of admission, lots of data are collected — some to provide sufficient information for the institution to complete registration in a programme and/or course, some to establish future study plans, some relating to the payment of fees, other information to allow the university to provide administrative support or additional other support if needed (such as if someone discloses a disability). These data would largely be contractual; some might be automated (fed in from existing study history held by the university), and some might also be volunteered by the learner. Some of these data will be attached to the learner's profile so that advisors and administrators can easily access it as needed. Once study begins, the learner starts to generate activity traces (websites visited, forum posts, etc.) which are mined by the system. Most of these data will be filtered in some way and stored on the university servers or perhaps fed back to the learning system. As Figure "An example of data journey in learning analytics" shows, information collected as part of admission, from trace data and a growing assessment record, is sent into a learning analytics system.

The example shows three ways in which learner data are used (there could also be others). The first, shown here in the form of a dashboard, is a means of sharing with the learner their progress. The second approach shows an alarm system that could notify the learner, or her academic advisor or teacher, to intervene. Such a system might be triggered by a missing assignment, a lack of online engagement, or predictive analytics. The third component suggests how the system might be adapted to fit the needs and preferences of the learner. The institution, teacher or advisor provides feedback to the learner, the system or the learning process for enhancements.

For more insight into the data journey, watch this short video.

Watch Video: https://youtu.be/LdkRch9yn24

Video Attribution: [Mapping the Data Journey in Learning Analytics]() by [Commonwealth of Learning]() is available under CC BY-SA 4.0.

Summary and Conclusion

We end this unit with the following key points:

- Data in education are valuable, precious and necessary commodities for learning analytics.
- Educational data vary in type, value and classification. Learners may give their data by self-reporting or through computer-based automated systems. Other types include contextual and directed data.
- Sources of data in the educational context have grown with the application of computer-based systems, artificial intelligence and machine learning.
- Institutions harvest data that exceed our knowledge. It is questionable whether all these data are used to optimise the learning experience of learners.
- Mapping the learner data journey requires exploration and careful thinking. You should identify the users (stakeholders), define the goals, draw and design, and finally refine and iterate.

We hope you enjoyed this lesson. In Unit 5, we explore uses of learning analytics.

Check your progress

a. One learner starts a discussion in a group of learners using the popular messenger service WhatsApp. The data the learner generates is an example of automated data.

   i. True
   ii. False

b. A primary school course that includes seven video lectures is an example of activity data.

   i. True
   ii. False

c. Samia is a student at a primary school. Her parents provided the school with her full name, gender and date of birth to complete the registration process. The provided data by Samia's parents are.

   i. gifted data
   ii. volunteered data
   iii. automated data
   iv. contractual data

d. Artificial intelligence and machine learning have introduced new types of data.

   i. True
   ii. False

e. Mining learner mouse clicks is an example of

   i. contextual data
   ii. consent data
   iii. log data
   iv. academic data

f. Some learner data are mined without linkage to learning optimisation.

   i. True
   ii. False

g. Digital exhausts can be considered

   i. data sources for computer-based educational platforms
   ii. feedback information to higher education
   iii. sources for mapping data journey in education
   iv. all of the above

h. Researchers, governments and decision makers are learning analytics stakeholders.

   i. True
   ii. False

i. Mobile platforms and wikis can be identified as learning sources for learning analytics.

   i. True
   ii. False

j. When mapping the data journey of a course you created, you may consider

   i. stakeholders
   ii. data types
   iii. data sources
   iv. all of the above

UNIT 5

Uses of Learning Analytics

Introduction

Welcome to the half-way mark in this course! We have come a long way — from starting with a very broad introduction to the world of learner data, to exploring notions of evidence and data. Unit 3 introduced academic and learning analytics and pointed to teacher/teaching analytics as an emerging phenomenon. Unit 4 moved on to exploring the sources of learner data.

While the title of this unit is "Uses of learning analytics," we have encountered various uses of learning analytics in all the units up to now, such as understanding learning and the learner's learning context (Unit 4), getting a sense of how learners are progressing toward achieving the envisaged outcomes of a course, and/or identifying opportunities to support learners who may face difficulties or need extra stimulation (Units 2 and 4).

In this unit, we explore different uses of learning analytics in more depth, and provide examples of how learning analytics is used to inform instructional design, pedagogy, assessment strategies and student support.

Learning outcomes

After you have worked through this unit, should be able to:

- evaluate the different uses of data analytics (descriptive, diagnostic, predictive, prescriptive, applied) and select two possible uses in your local institution or courses
- write a draft article for your local institution staff or student newspaper in which you provide a rationale for the use of learning analytics and illustrate the potential by referring to two specific uses in your course/institution

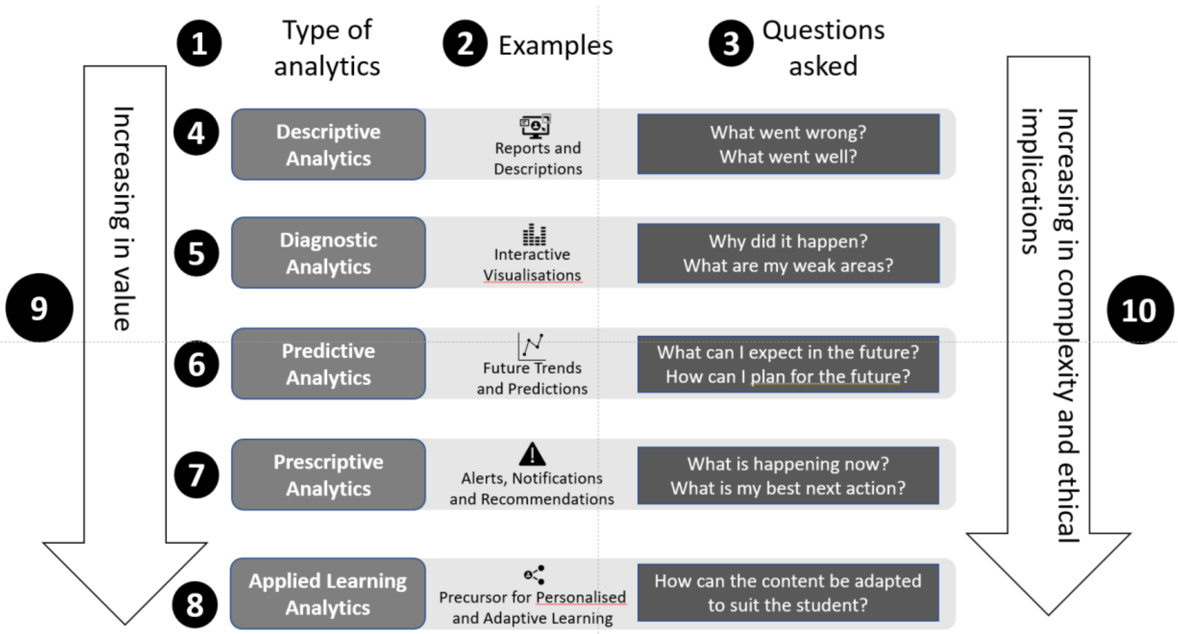
At this stage, please watch the video "Uses of learning analytics."

Watch Video: https://youtu.be/2whbbGcDeJ8

Video attribution: "Unit 5: Uses of Learning Analytics" by Commonwealth of Learning is available under CC BY-SA 4.0.

Uses of learning analytics: an overview

Another way to look at the uses of learning analytics is shown in figure "An overview of different types of analytics".[16]



An overview of different types of analytics

You will notice that the figure includes numbering aimed at helping us to walk through each part. The main threads of the illustration are “type of analytics” (1), examples (2), and questions asked (3). You will also notice two arrows (9 and 10) going down each side of the main illustration. We will return to these later.

There are five suggested types of learning analytics: descriptive (4), diagnostic (5), predictive (6), prescriptive (7), and applied (8). These are described further in the following sections.

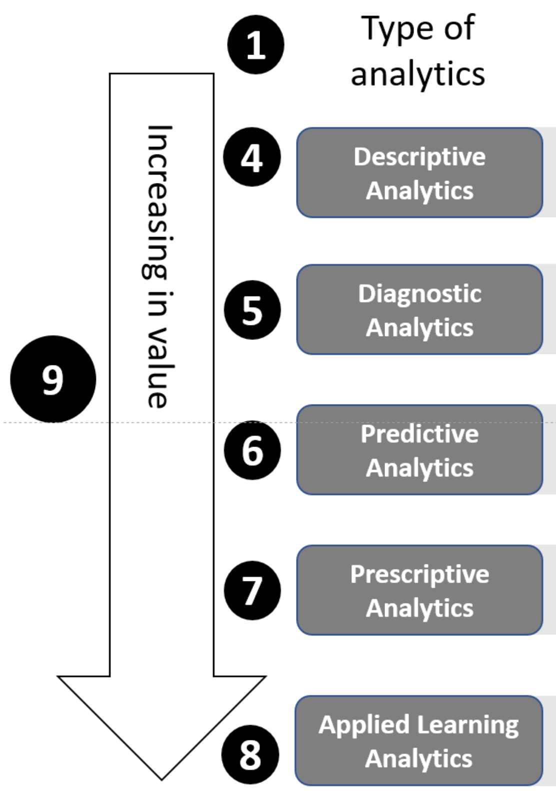
[16] Adapted from https://commons.wikimedia.org/wiki/File:Types-of-Learning-Analytics.png.

Descriptive analytics

Descriptive analytics asks questions to help us understand what went well or not so well. Examples of descriptive analytics include reports and descriptions. So far in this course, we have mentioned several examples of descriptive analytics, such as the number of learners who passed an examination sitting, how technology has impacts on teaching and learning, and so forth. The data collected (qualitative and/or quantitative) are analysed to *describe* trends and patterns. It is important to note that descriptive statistics do not try to explain "why" something happened or "why" a trend is emerging; it just describes the trend or phenomenon itself.

This is a time of data revolution, with organisations having access to not only greater volumes of data, but also a greater variety and granularity of data, often in real time. Some scholars and researchers claim we no longer need to *understand* what is happening, as long as we can *see* what is happening or has happened, and respond as quickly as possible. For big corporations, their sustainability and profitability can depend on how quickly and appropriately they are able to respond to emerging trends. Some scholars state that in a context where time is of the essence, responding is more important than understanding. Do you agree?

While we understand there may be situations or contexts where responding as well as one can may be more appropriate than understanding why something is happening, these situations and/or contexts should be outliers rather than the norm. A reasonable example of responding to something happening without necessarily understanding the reasons behind it is when you notice someone drowning and there is just no time to ask why they're drowning. Determining the why can be established later when, the person is safe.



An overview of different types of analytics

In the context of education, many institutions may be tempted to respond to a phenomenon without first trying to establish why a pattern is emerging or why, for example, learners have failed a particular course. We can surely all recall examples where a manager or head of department demands that teachers or support staff report how they will prevent this from happening again. It comes as no surprise that often, these strategies and interventions are doomed to fail, due to hasty implementation without first establishing the reasons for the occurrence.

In the context of learning analytics, though descriptive analytics can help teachers and learners make more informed and appropriate decisions, diagnostic analytics add more value and can have a greater impact, as illustrated by the downward arrow (9) in the figure "An overview of different types of analytics".

Diagnostic analytics

Diagnostic analytics, unlike descriptive analytics, aim to determine the reasons *why*, for example, some learners fail a course more than once, or to uncover the factors (or variables) that have an impact on learners' performance in general, or on specific courses. Once researchers have identified the underlying reasons, they are better able to design an appropriate response — e.g., a dynamic, interactive dashboard allowing teachers and learners to make more informed and appropriate decisions.

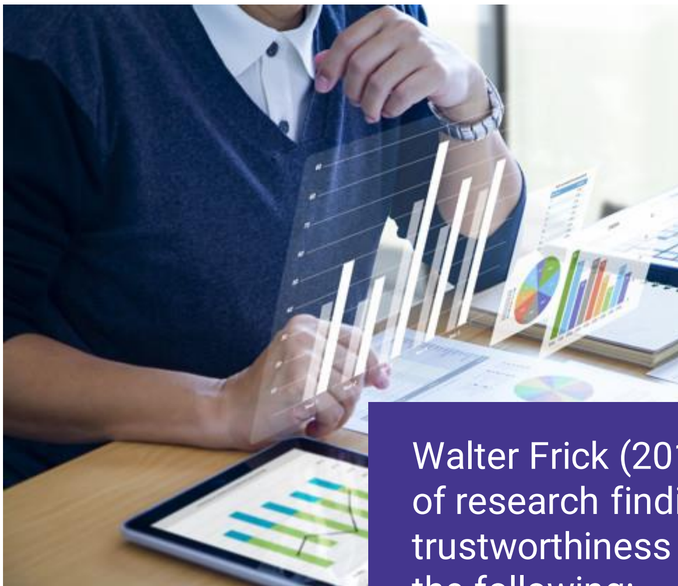
It is important to note that both descriptive and diagnostic analytics are either backwards looking or dealing with what is currently happening or emerging. In a certain sense, we could say that both these types of analytics are providing information in hindsight.

Predictive analytics

Predictive analytics, unlike descriptive and diagnostic analytics, focuses on the *future*. The questions informing predictive analytics include but are not limited to "What can we expect to happen in the future?" and "How can we plan for the future?"

While prediction is fascinating, we can all probably think of examples where something was expected to happen but never did, or happened, but not in the way it was predicted.[17] The likelihood of a specific prediction coming to realisation depends on many things, such as the expertise of the person making the prediction, and the data on which that prediction is based.

[17] Two fascinating books by Nassim Nicholas Taleb illustrating the surprises that often await those relying on predictions are Fooled by randomness: The hidden role of chance in life and in the markets (Random House, 2005) and The black swan: The impact of the highly improbable (Random House, 2007).



Walter Frick (2015) [18] discusses a range of research findings about the trustworthiness of research and mentions the following:

- the intelligence of the person making the claim
- researchers' expertise in the field in which they are making predictions
- the experience in the field — "practice improves accuracy"
- "teams consistently outperform individuals"
- open-minded individuals make better predictions
- training in probability guards against bias
- "rushing produces bad predictions"
- revising data, analysis and the forecast or prediction produces better results

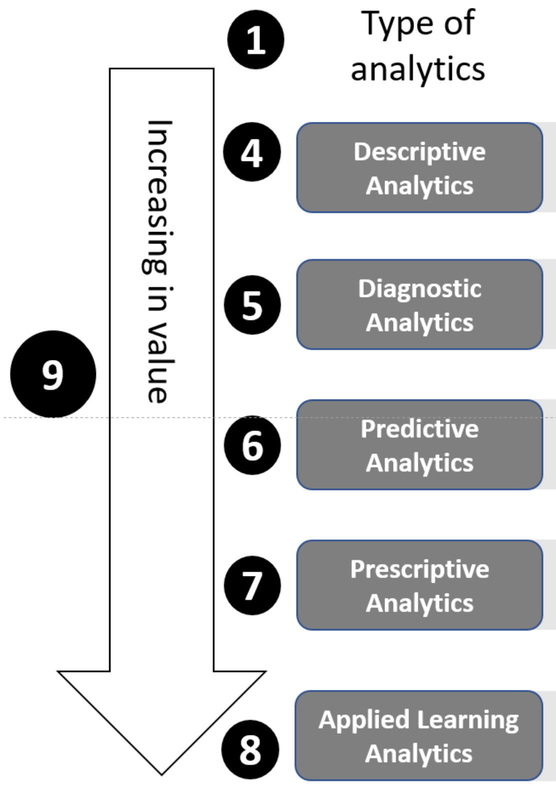
[18] Frick, W. (2015, February 12). What research tells us about making accurate predictions. Harvard Business Review. Retrieved from https://hbr.org/2015/02/what-research-tells-us-about-making-accurate-predictions

Diagnostic analytics

Though the above list is interesting, it is also important to link this discussion about predictive analytics and its trustworthiness to our earlier discussions in Unit 2 on evidence and data, and specifically to the criteria for scientific reliability in quantitative research, and trustworthiness in qualitative research.

There are two particular issues to flag for predictive analytics: the responsibilities and care needed when using predictive analytics in education, and the complexity of predictive analytics when we consider education as an open and recursive system, also discussed in Unit 2.

Increasingly, the findings and analysis resulting from predictive analytics are communicated to teachers and learners through interactive dashboards. For example, teachers may have access to a dashboard showing how learners are progressing through the course, whether as individuals or as a class. The visualisation may include, amongst other aspects, class attendance and/or login data, engagement data, results from formative assessment opportunities, and other recorded data. So far, this sounds like descriptive analytics, and many dashboards are no more than this.



An overview of different types of analytics

As the downward arrow (9) in the figure "An overview of different types of analytics" indicates, if a dashboard “stops” at descriptive analytics, the value added to learning may be less than if the dashboard also employs some element of predictive analytics.

For example, imagine a dashboard that includes predictive outcomes. The software might use aspects of known information (e.g., looking at how previous learners have performed on the course, combined with data about the individual learners) to provide an indication of the learner’s probability of passing/failing the course. When a teacher has access to this information, she can respond by providing additional support or refer the learner to other sources of help. Such predictive insights are potentially very valuable and could prompt much needed action.

However, predictive analytics requires a particular skill set, experience, familiarity with statistics and software, etc., and it should not be attempted by unqualified individuals. As well as the inherent complexities, the ethical implications are also potentially significant and long lasting.

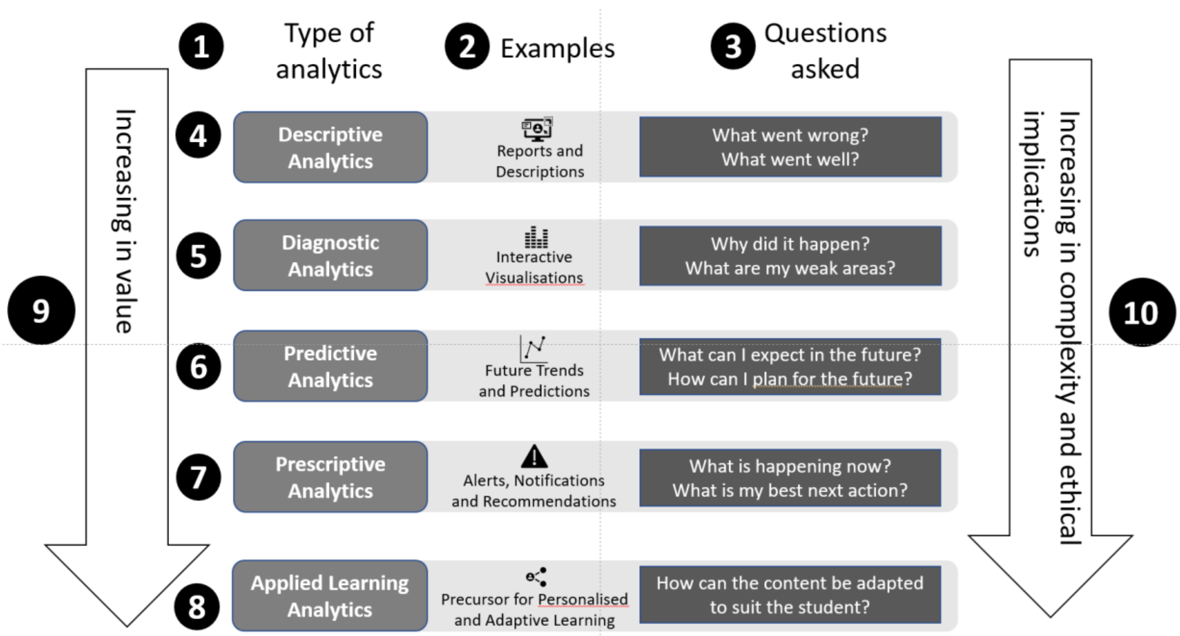
If an event is described incorrectly in a descriptive context, the implications are easier to remedy and less likely to cause disruption than, for example, a prediction that learners with characteristics x or y will pass or fail. Earlier, we highlighted the benefits of having such predictive analytics presented to teachers and individual learners.

Now, consider the situation if the data on which the predictive analytics was based were biased or incomplete, or if an important variable on which the predictions were built had changed since the data were collected (refer back to the discussions in Unit 2 on causality and correlation, if needed).

Many scenarios can unfold when predictive analytics are misleading, such as wasted time and resources, negative psychological impacts on learners, and the possibility that the label “at-risk learner” might accompany a learner for the rest of his years at the institution. Hence, there is a balance to be struck between the potential benefits and the potential harms of predictive analytics.

Prescriptive analytics

Prescriptive analytics moves analytics to *how* to realise a particular future. Descriptive and diagnostic analytics deal with the past and the present, and predictive analytics suggests what might happen in future if the current trend or pattern continues. Prescriptive analytics works with predictive analytics to prescribe what is needed to achieve a particular outcome. Here is an example. Research at a particular institution has shown that learners who do not have Mathematics at a particular level have previously been linked to fail results for Economics and Accounting. Based on this, a prediction is made that your chances of passing Economics or Accounting without Mathematics at a particular level are fairly low. When the academic management of the institution looks at these analytics, they make a decision to require (*prescribe*) learners without Mathematics to first complete a bridging course in Mathematics before they can enrol in Economics and/or Accounting.



An overview of different types of analytics

Thinking back to the downward arrows (9 and 10) in the figure "An overview of different types of analytics", it seems clear that prescriptive analytics may add even more value than predictive analytics, but also that prescriptive analytics bears a greater ethical burden. In predictive analytics, there is the possibility that a learner may prove the analytics wrong. In prescriptive analytics, learners who may have proven the analytics wrong would be prevented from doing so exactly because they have taken a particular path (or been prevented from taking a path), based on the prescribed steps.

This example helps to illustrate that the responsibility associated with prescribing certain actions to learners requires both diligence and expertise. In addition, learners and other stakeholders should understand the basis for the prescription.

Applied analytics

This final category is in many respects the culmination of the previous four forms of analytics. Applied analytics comes into play when decisions from the other forms of analytics are used to personalise learners' learning journeys, often in real time.

Consider the following scenario. Research has found that learners who study in a second language have a greater probability of falling behind in a course than learners who study in their home language. Of course, language is not the *only* variable, but yes, research shows a strong correlation between studying in a foreign language and falling behind. (Recall the discussion in Unit 2 on correlation and causation.)

- Ishmael (not his real name) has enrolled in a 12-week online History course. There are 100 other learners in the course, and an instructional team (one dedicated instructor and ten contract learner support staff who work with the instructor). Before the course "opens" and instruction starts, the instructor and support team go through the list of learners and see that Ishmael has been flagged as studying in a second language. The online registration system automatically flags such learners (a result of previous years' descriptive, diagnostic and predictive analytics). Ismael is one of 40 learners who meet this criterion. The instructional team activates an alert system to warn the team if Ishmael fails to log in for two days, does not submit an assignment, does not participate in the online discussion forums, and/or has difficulty passing the automated self-assessments.

- The course begins, and the online platform is opened for learners to introduce themselves in the discussion forum, download materials and access the sequenced learning units. The system[19] alerts the instructional team that Ishmael has attempted to log in but seems to be having trouble. The system is set up to also send an automated short text message to Ishmael's mobile number, providing guidance. The system sends the instructional team a notice to indicate that the message was delivered and opened, and a second notice to indicate that Ishmael has successfully logged on to the course platform. A member of the instructional team sends Ishmael a personalised message welcoming him to the course and assuring him that the team will support him wherever they can. Over the next three days, the team watches the performance of Ishmael closely to get a sense of his progress.

[19] We acknowledge that speaking of "the system" sounds as if it runs automatically and acts autonomously without any human input or oversight. Of course, this is far from the truth. In referring to the system, we acknowledge that whatever system has the ability to respond to learners' behaviours has been programmed by humans. Recent advances in machine learning and artificial intelligence have created the potential for adaptive LMSs to run autonomously once they have been set up. See Unit 10 for a discussion of the potential, challenges and risks of human algorithmic collaboration in supporting learners.

- The system alerts the instructional team that Ishmael has trouble passing the first self-assessment exercise — this consists of a video that learners watch before attempting ten multiple choice questions. Looking at the data trail of his engagement, the team notices that Ishmael never finishes watching the video but (unsuccessfully) tries to answer the questions. They activate an email to him with further explanation of the role of instructional videos and advise him to first watch the video in its entirety before attempting the self-assessment questions again. The system alerts the team that Ishmael finished watching the video and has got most of the self-assessment questions correct, except for one particular question. Based on the subject matter of this question, the system diverts Ishmael's learning trajectory to web pages that specifically deal with the subject matter that he has yet to understand and apply. The web pages contain a self-assessment exercise allowing Ishmael to test his comprehension. This time around, he succeeds with flying colours, and his trajectory rejoins the main course outline. The instructional team is alerted that Ishmael has now acquired the necessary competency that he lacked.

Moving from data to wisdom

Learner data, whether digital or non-digital, can provide teachers with crucial information on how to teach more effectively, appropriately and ethically. Some analytics may be more useful than others but may also require a particular skill set, experience and expertise. Each potential user of learning analytics should consider their own skill set as well as any expertise available within their institution.

Unit 6

In Unit 6, you will explore various methods and tools that you can use to analyse data.

Unit 7

Unit 7 will then introduce you to the ethics of the collection, analysis and use of learner data, and explore issues of privacy and consent. In the excitement of all the possibilities as well as the various tools and methods, it is important not to lose focus of what learning analytics is about: understanding and supporting learning and the contexts in which learning occurs.

Unit 10

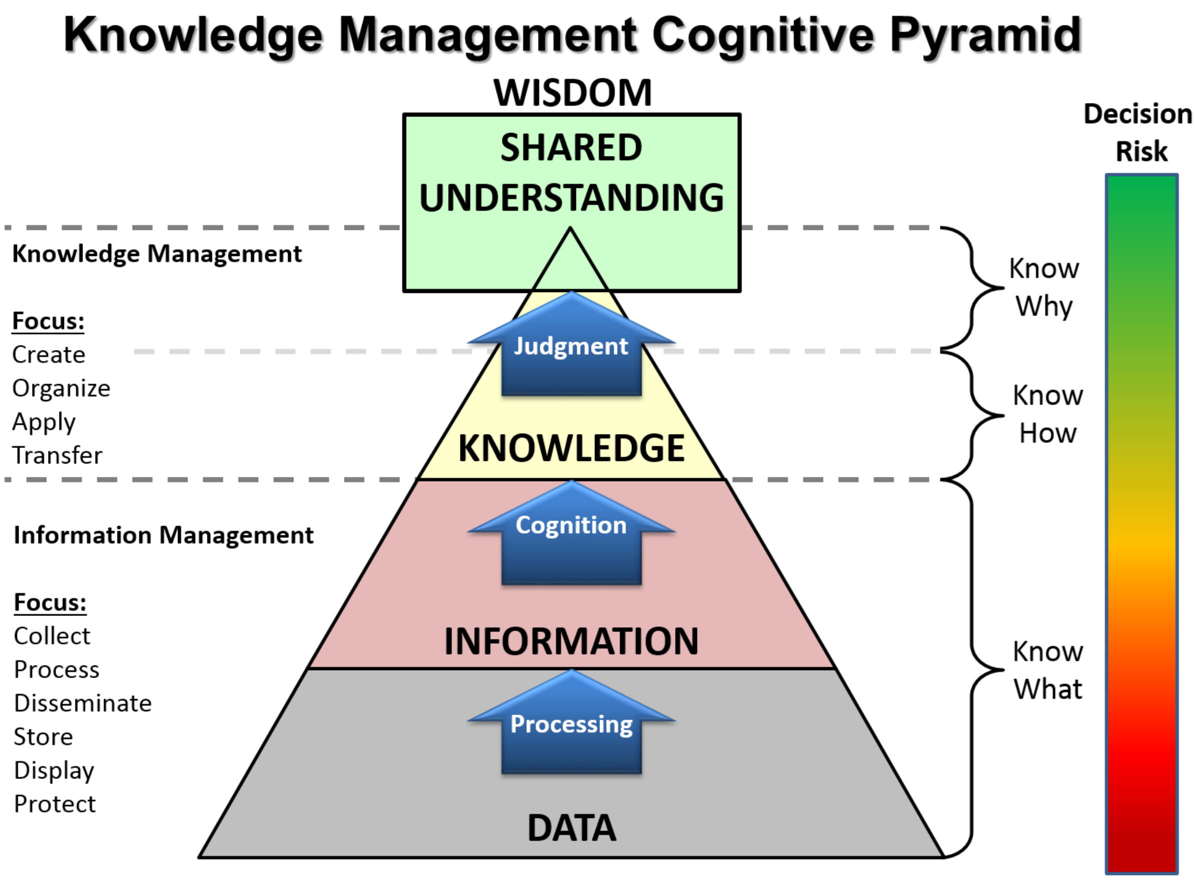
In the last unit of the course, we will respond to criticism and concerns that the increasing collection, analysis and use of learner data are not justified and resemble carceral surveillance.

As authors who recognise issues pertaining to ethics, privacy and the dangers of surveillance, we also believe that learning analytics can be an ethical and caring practice. We further believe that it can help institutions fulfil their fiduciary duty to society and learners. We have included a list of useful references and resources at the end of the course, should you like to read more.

In line with our own position on the potential and challenges in learning analytics, we conclude this unit with some thoughts on moving from data, to information, to wisdom.

- Various authors have pointed out that “data precedes information, which precedes knowledge, which precedes understanding and wisdom.” [20] As data moves to become wisdom, it goes through several iterations: reduction, abstraction, process, organisation, analysis, interpretation and application.
- When learners’ data are collected, analysed and produced — whether in reports, visualisations or adaptive learning — the purpose is to lead to ethical and appropriate action. While we can think of many examples where an action or intervention was not based on sound data or analysis, the aim of learning analytics is to collect and use learner data to result in wisdom. Figure "The knowledge management cognitive pyramid" illustrates the move from data to wisdom, and the different stages and uses.

[20] Page 9 in R. Kitchin. (2014). *The data revolution*. Sage.



The knowledge management cognitive pyramid[21]

On the left of the pyramid is a list of the stages that move data toward information – ranging from collection and processing to disseminating, storing, displaying and protecting. Moving from data to information resembles, to a certain extent, descriptive and diagnostic analytics on the one hand, and predictive and prescriptive analytics on the other hand. Descriptive analytics provides an overview of *what* is happening, and diagnostic analytics provides the answers to *why* something is happening. When there is no response to this information, it stays as data and information. The Knowledge Management Cognitive Pyramid proposes that when data and information become knowledge, this process involves acting on the information and includes steps such as creating, organising, applying and transferring. Wisdom as the end result of this process means a shared and evolving understanding.

[20] Retrieved from https://commons.wikimedia.org/wiki/File:KM_Pyramid_Adaptation.png.

What is particularly interesting about Figure "The knowledge management cognitive pyramid" is the changing colour of the bar on the right, representing the risks associated with moving toward wisdom. If we collect data that are biased, incomplete or based on ill-informed decisions, the risks to the quality of the outputs are high and likely to directly impact whether the final result emerges as *meaningful* insight.

Reflection action: making progress

The two outcomes envisaged for this unit are as follows:

- Critically evaluate the different uses of data analytics (descriptive, diagnostic, predictive, prescriptive and applied), and select two possible uses in your local institution or courses they teach.
- Write a draft article to your local institution’s staff or student newspaper in which you provide a rationale for the use of learning analytics and illustrate the potential by referring to two specific uses in your course/institution.

Getting practical

Look at the different types of data analytics in the context of the collection, analysis and use of student data, and critically evaluate each of them — the benefits, and also the risks, challenges and/or shortcomings. Using below given table on the next page may actually provide you with a very useful tool for future use!

Type of data analytics	What are the distinctive characteristics of this type of analytics?	What are the promises and/or benefits of this type of learning analytics?	What are the risks, challenges and/or shortcomings in using this type of data analytics?
Descriptive analytics			
Diagnostic analytics			

Predictive analytics			
Prescriptive analytics			
Applied analytics			

Critically analysing the types of data analytics

Making the case for learning analytics

Draft an article to your local institution's staff or student newspaper (or imagine that you have one, if needed), in which you provide a rationale for the use of learning analytics and illustrate the potential by referring to two specific uses in your course/institution.

We suggest you consider using the following sections to structure your article:

- Introduction: Briefly refer to how you, in your teaching practice or in the institution, have relied on learner data to make more informed decisions about pedagogy and learner support.
- Introduce them to learning analytics: It is important to know your audience and write this section so that your audience — whether they are colleagues and management, or learners — will find it interesting. For example, you can refer them to your own experience of wanting to teach more effectively, and learning analytics providing you with the potential tools to do so. Or if you are writing to the student newspaper, link learning analytics to their need for more understanding and more appropriate support from teachers and the institution.
- Give your audience a view of the potential of any of the above uses of analytics in your courses.
- Conclude your article with an invitation to those who are interested to also do this course or to make contact with you.

Summary and conclusion

This was a bulky but interesting unit where we tried to get a sense of how learning analytics might play out in an educational institution. Though earlier units hinted at this, Unit 5 has sought to provide a more detailed mapping of the different forms of data analytics.

The final activity encouraged you to write a personal, critical overview of the five types of data analytics as well as to think about how you will justify the use of learning analytics in your course and/or institutional context. Carrying out such an exercise really helps you to understand what implementation might actually look like for you.

We conclude this unit by recapping the final discussions: the link between data, information, knowledge and wisdom. We have no doubt that learning analytics, since its inception, is moving toward wisdom — applying data, information and knowledge to assist learners and teachers to make better decisions. While the complexity and ethical implications increase as we move from descriptive analytics to applied analytics, it is also important to realise that the basis for all these forms of data analytics, and the move from information and knowledge to wisdom, is *data*.

Check your progress

We sincerely hope you enjoyed discovering the different types of analytics and getting a sense of how learning analytics unfolds, from using data to describe what students are experiencing or what is happening, to actually shaping their learning journeys as they progress through a course using applied analytics. Use the following questions to map your own progress.

a. Look at the questions in the left-hand column, and match these questions with the type of analytics that will help you answer them.

Question	**Answer**	**Which type of analytics will answer the questions on the left?**
How can the content adapt to suit the learner?		Descriptive analytics Diagnostic analytics Predictive analytics Prescriptive analytics Applied analytics
What went wrong? What went well?		
What is happening now? What is my best action?		

Why did it happen? What are my weak areas?		
What can we expect in future? How can I plan for the future?		

b. Look at the different types of analytics in the left-hand column, and match these types with examples of what they may look like in everyday practice.

Type of analytics	**Answer**	**Examples of what the end result of this type of analytics may look like**
Descriptive analytics		Precursor for personalised and adaptive learning (as illustrated in the example of Ishmael) Future trends and predictions Reports and descriptions Alerts, notifications and recommendations Interactive visualisations
Diagnostic analytics		
Predictive analytics		
Prescriptive analytics		
Applied analytics		

c. Which one of the five types of analytics will you use when you have the following questions?

Example of questions asked or uses of learning analytics	Answer	Which type of analytics best addresses the questions on the left?
I want to provide the learners in my class with a visualisation that will help them to understand their progress and improve their chances of success.		Descriptive analytics Diagnostic analytics Predictive analytics Prescriptive analytics Applied analytics
I am not sure what is happening in my online class. I see some of the learners are not participating, but I am not sure who is not participating.		
Based on research findings, I have a good idea which of the learners in my class are at risk of failing. I want to send them personalised emails based on their progress and provide them with additional support based on their personal profile.		
When I look at data of my learners' prior learning experiences, their home language and their previous marks in Mathematics, I want to know which may be at risk of failing Accounting.		
Our LMS allows us to adapt each learner's journey according to their profile, previous learning experiences, and learning needs. What type of analytics can I use to actually shape each learner's journey according to their profile?		

d. Predictive analytics adds more value to our understanding of learners' learning than descriptive analytics.

   i. True
   ii. False

e. Diagnostic analytics is more complex than applied learning analytics.

   i. contextual data
   ii. consent data
   iii. log data
   iv. academic data

f. Collecting, analysing and using learner data to help learners achieve their learning outcomes must be done ethically. Prescriptive analytics have greater ethical implications than diagnostic analytics.

   i. True
   ii. False

UNIT 6

Learning Analytics: Methods & Working with Data

Introduction

When we talk about learning analytics methods, these include a range of methodologies and techniques inherent in educational data mining and statistical analysis — from basic descriptive statistics to structural equation modelling, the use of machine learning and, increasingly, artificial intelligence.

The definition of learning analytics carries three key parts, which constitute methods:

i. collection,

ii. analysis and

iii. reporting.

In this unit, we will discuss each of these parts, and introduce you to techniques, methods and tools that will help start your practical learning-analytics journey.

Learning outcomes

The course of this unit will follow the route one might take when exploring an educational data problem and conducting analysis in that context. After you have worked through this unit, you should be able to:

- identify the different analytical and statistical methods that are most often used in learning analytics, including but not limited to text analysis and process mining
- describe the uses of machine learning and artificial intelligence in learning analytics
- choose a context-appropriate option to make the optimum use of algorithmic decision-making systems (and understand the implications)

Data collection

One key task in learning analytics is to know what data to collect, as well as where and how to collect data. Learning analytics involves qualitative and quantitative data collection methods. In this part of Unit 6, we focus on collection methods for quantitative data. (You will recall from Unit 2 that qualitative data collection includes structured and unstructured interviews, focus groups, observations, open-ended surveys and other narrative sources; if you need a refresher, you can review section "Making sense of knowledge claims".)

The quantitative approach (an approach for numerical and/or measurable data) in learning analytics tends to give rise to large data sets. For example, when learners use digital learning environments, they generate lots of data that are stored on servers. This can lead to extremely large repositories of data sets, sometimes known as "big data."

Data manipulation

Data manipulation is critical when validating your quantitative sample. In the data science world, we say “no clean data, no clean mining results.” If the data source is a learning management system, a digital online platform or similar, it is highly likely that your hands will be dirty with rogue data. Data file extensions may be comma delimited, tab delimited, or pipe (|) delimited. As a learning analyst, you will almost certainly reach the stage where you work with a messy information structure. Log data from digital educational systems are most likely not immediately amenable to analysis (Slater et al., 2017). Hence, you must address your repository data types, values and rows.

Data cleaning milestones

Noisy data are not intentional but incidental, arising due to technological and human flaws. In the following, we list some of the major cleaning steps needed when working with educational data for the purpose of implementing learning analytics techniques/processes:

Identifying missing values

As log files sometimes contain undocumented values for a variable, you might need to consider imputing missing values (e.g., using mean or median values for numerical data or imputing some sort of random input for non-numerical variables).

Removing duplicates

Identifying and removing identical records.

Removing redundant data columns

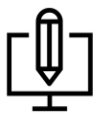
Many educational data repositories will include redundant information. Erasing these data will make your analysis cleaner and faster.

Dealing with data types

Validate that your records of numerical data belong to a numeric class, letters belong to strings, base numeral belongs to binaries, etc.

Marking outliers

At some stage of the analysis, you may encounter abnormal values that exert a disproportionate influence on your results. You might consider deleting, transforming or assigning new values to data outliers.



Correcting inconsistent data

Values for data variables may not be consistent and may not follow one standard. If you have a data set with varying formats or other inconsistencies, first standardise or harmonise it to create a cohesive data set.

Entry and intermediate level

If you are new to data manipulation or have a little experience, below table presents our software recommendations for tasks of simple to medium complexity:

Tool	Description
Microsoft Excel	Machine or cloud spreadsheet software from Microsoft
Google Sheets	Cloud spreadsheet software from Google
Zoho Sheets	A cloud spreadsheet tool
LibreOffice	A free, open-source machine-based spreadsheet tool
Tableau	Focuses on front-end visualisation but can handle simple data manipulation tasks

Entry- and intermediate-level tools for data cleaning

Advanced level

Spreadsheet software will not be sufficient for all types of data manipulation tasks. You may reach the point where there is a need to do advanced data aggregations, creating a more complex task not easily tackled by entry-level software. Such basic tools have limitations associated with the amount of data being pre-loaded, loaded, and executed, which slow down the analysis process. In given table, we suggest a few more advanced data-manipulation tools that you might use for both simple and complex tasks.

Tool	Description
SPSS	Statistical software platform provided by IBM
SAS	Data management and advanced analytics tool
R	Free software environment for statistical computing and graphics generation
STATA	Statistical software for data science
Python/Java/C++	Powerful programming languages that can be used for all sorts of data manipulations

Advanced-level tools for data cleaning

Learning analytics techniques

Having discussed data collection and manipulation for learning analytics, we now move on to discussing learning analytics techniques. There is a vast selection of popular methods and techniques for implementing learning analytics in order to understand, explain and solve educational applications and problems. Such techniques originate from the broader fields of machine learning, data mining, visualisations and statistics. Below, we provide a selection of some of the more popular techniques and methods for the analysis stage, taken from the literature (Leitner et al., 2017; Romero & Ventura, 2020).

Prediction

As noted in Unit 5, this is one of the most widely used techniques in the field of learning analytics. Prediction is used to forecast student performance, identify student behaviours, warn instructors and empower course designers.

Clustering

This is a key part of the science of data mining. Clustering brings together similar materials or learners based on their learning and interaction patterns. There are many examples of how to use clustering (Bharara et al., 2018), among them: to identify learning performance assessment rules; to provide personalised eLearning environments based on learner personality; to classify the eLearning behaviour of learners; to group learners on the basis of their cognitive flairs; and to recommend the best course combination for individual learners.

Outlier detection

This is used to indicate significantly dissimilar individuals/learners — for example, this may be used to detect learners with difficulties or unusual learning patterns. However, this method should be used with caution, as it can be associated with inadvertent labelling and other ethical consequences.

Relationship mining

Part of the data mining field, relationship mining aims to identify links between variables, but in a complex way. That is, it digs deeper to understand which variable is most closely linked to other variables. An example of relationship mining in learning analytics is identifying relationships in learner behaviour patterns to diagnose student difficulties.

Social network analysis

This seeks to understand a community by mapping the relationships that connect the community's individuals as a network, and then trying to draw out key individuals, groups within the network ("components"), and/or associations between individuals. Often abbreviated as SNA and used to examine collaborative learning environments by visualising relationships between learners, it can be powerful for identifying isolated learners.

Process mining

Also known as PM, this is a method to extract process-related knowledge from a log of events. One of the most popular types of PM is sequential pattern mining, which seeks to find relationships between occurrences of sequential events, such as the individual tracing of multiple actions in grocery stores. Examples of process mining in education are exploring and visualising detailed traces, delivering a set of complex grade sequences, and examining students' social interactions and timestamps of their behaviour over a long period of time.

Text mining

Text analysis is increasingly important in education. Several sources are available for exploring text learning patterns, including discussion forums, assignments, essays, chat, documents, and web pages. Text mining and analysis can be used to examine student support and motivation, recommend courses, conduct sentiment analysis, and generate automatic content and questions.

Distillation of data for human judgment

The main objective of this technique is to outline information in a visual way to aid decision making. In learning analytics, it has been used to help instructors to visualise and analyse the ongoing activities of learners and to provide information related to learning and course design.

Discovery with models

This is very much connected with prediction, clustering and other data mining methods. Some examples of how it can be used in education include identifying relationships among student behaviours and characteristics or contextual variables, and integrating psychometric modelling frameworks into machine-learning models. Discovery with models leverages the examination of conceptual frameworks that are brought into empirical work.

Gamification

This facilitates playful learning to maintain motivation — e.g., the integration of achievements, experience points or badges as indicators of success. Learning analytics can take advantage of gamification to enhance learning through the provision of more exciting representations of data and the use of incentives.

Multimodal analytics

Multimodal learning is concerned with multiple data sources, including the senses — visual, auditory and kinaesthetic. In learning analytics, it applies machine learning methods and sensor technologies (e.g., heartbeat, virtual reality, augmented reality. etc.) to present and explain new learning insights.

Machine learning (ML) and artificial intelligence (AI)

These are strongly related fields utilised to advance digital and smart education systems. ML methods with AI require advanced skills, such as using Naive Bayes, regression models and decision trees to provide effective advice, actions and perspective analytics. Some ML and AI practices seek to find insights in data automatically using intelligent models that are exposed to new data and adapt independently.

Descriptive statistics

This helps us understand the characteristics of data sets. You will no doubt be familiar with some types of descriptive statistics, such as the average (mean), mode and median. Other important ones are percentiles, percentages, count, and standard deviation.

Inferential statistics

This involves more powerful and complex methods than descriptive statistics. You might use inferential statistics when descriptive methods are limited. Inferential statistics allow you to make predictions ("inferences") from data taken from samples by making generalisations about the population. Methods include t-test, regression analysis, analysis of variance (ANOVA), analysis of covariance (ANCOVA), and chi-square.

If you are interested in learning more about these techniques, follow the open education resource links in the footnotes.[22] [23]

[22] https://sgfin.github.io/learning-resources/#cheatsheets

[23] https://www.solaresearch.org/wp-content/uploads/2017/05/hla17.pdf

Learning analytics reporting: visualisations

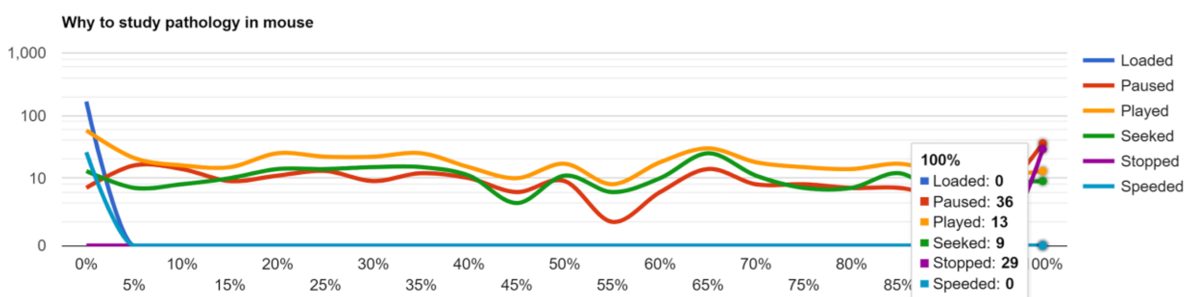
We now come to the final stage of learning analytics methods: reporting. The role of reporting is usually to give a summary to learners, instructors, researchers, course designers, decision makers or higher management. It is most often written (i.e., reports, best practices, fact sheets) or visual. Visualisations can be particularly useful; for example, learners can view their progress in assignments and classes, instructors can get an overview of their courses, and decision makers can be supported in making financial decisions.

Visualisation tools

Visualisations in learning analytics effectively translate numbers into images. There are numerous approaches, such as charts, flowcharts, mind maps, 3D plots, scatterplots, pie charts and infographics. Recent research suggests that dashboards have become increasingly important, making outputs easy to understand, providing better visibility and offering informative insights. In below table, we list tools that can be used to generate either static or interactive visualisations of your data analysis.

Tool	Description	Skills level
Spreadsheets (Excel and Google)	Create static charts that provide a descriptive explanation of your analysis.	Entry and Intermediate
Tableau	Build rich and interactive dashboards.	Entry and Intermediate
Javascript (D3 and Google Charts)	Build complex and interactive data visualisations that can be used on web browsers (see Figure below "Google Chart interactive figure displays learners' engagement with a MOOC video")	Intermediate and Advanced
R (package: ggplot)	A popular package that enhances R's potential for data visualisation.	Intermediate and Advanced
Gephi	Social network analysis reporting and analysis tool	Entry and Intermediate

Tools for data visualisation



Google Chart interactive figure displays learners' engagement with a MOOC video

(From Khalil & Belokrys, 2020.)

All-purpose learning analytics tools

Many types of software are available for learning analytics, including off-the-shelf tools. In most instances, you will need to employ more than one tool — for example, dedicated software for creating dashboards. A selection of learning analytics tools is presented in Table in the next page.

Tool	Description
SNAPP[24]	Allows users to visualise networks and conduct social network analysis from social forums
Weka[25]	A collection of visualisation tools and algorithms for data analysis and predictive modelling, with graphical user interfaces
Rapidminer[26]	Advanced machine learning tool enabling data mining, text mining and predictive analytics
R, SPSS	Packages for interactive, or batched, statistical analysis
GIZMO[27]	Provides visualisation of learner engagement in online courses to tutors

All-purpose learning analytics tools

[24] https://github.com/sandeepmjay/SNAPPSakai-Beta/blob/master/INSTALL.md

[25] https://www.cs.waikato.ac.nz/ml/weka/

[26] https://rapidminer.com/

[27] https://gizmo-vis.github.io/gizmo/notationEditor/index.html

Summary and conclusion

We end this unit with the following key points:

- Learning analytics is a relatively young field, yet there are already several established approaches to gain insight and enhance the learning experience.
- The key stages in learning analytics methods are what, where and how to collect, analyse and report data.
- There are many different types and sources of data. You should aim to start your learning analytics exploration with data manipulation and cleaning.
- Analytic techniques are very much influenced by computer science methods; examples include machine learning and AI, data mining, and descriptive and inferential statistics.
- Learning analytics includes both quantitative and qualitative methods, but there is a greater focus on quantitative methods.
- A combination of tools and analytic approaches will enhance your uses of learning analytics more than a one-size-fits-all approach.

We hope you enjoyed this unit. In the next unit, we change tack and start to consider some of the ethical issues in the world of learning analytics.

Check your progress

a. Learning analytics includes quantitative and qualitative methods for doing data collection and analysis.

   i. True
   ii. False

b. In learning analytics, it is compulsory to standardise your data types and values so that your data set is harmonised.

   i. True
   ii. False

c. Learning analytics includes techniques and methodologies in:

   i. data analysis
   ii. measurement
   iii. data collection
   iv. all of the above

d. Learning management systems and MOOCs are

   i. eLearning platforms used by decision makers to develop insights on teachers
   ii. learning platforms that are rich in educational data
   iii. learning analytics quantitative methods
   iv. techniques forming part of machine learning and artificial intelligence

e. Chi-square, ANOVA, and t-tests are methods used to collect data about learners.

   i. True
   ii. False

f. Spreadsheet programs are adequate tools for simple and complex data analytics tasks.

   i. True
   ii. False

g. Which of the following can be considered a visual reporting tool for learning analytics?

   i. spreadsheets
   ii. SPSS
   iii. Google Charts

h. Removing duplicate records from educational data sets falls under:

   i. data analysis
   ii. data reporting
   iii. data collection
   iv. data manipulation

i. Machine learning and data mining sciences are methods of learning analytics.

   i. True
   ii. False

UNIT 7

Ethics, Privacy and Consent

Introduction

The application of learning analytics has huge potential to benefit both learners and institutions. Applied responsibly, targeted interventions based on observed behaviours or on predictive analytics, for example, can provide relevant and personalised prompts and improve learner outcomes. Core to the collection, analysis and use of data associated with learning analytics is a range of institutional moral and ethical duties, including learner privacy and consent issues.

In this unit, we explore some of the relevant issues, highlighting potential areas for focus and directing you toward relevant documents within your own institutions and national contexts.

Learning outcomes

After you have worked through this unit, you should be able to:

- explain the broad range of ethical issues associated with learning analytics
- critically assess the terminologies used, types of data collected, how that data might be used and who the data could be shared with, as indicated in the terms and conditions of different institutional websites
- examine legislation and/or policy frameworks to define and identify a range of elements such as "legitimate use," "personal data," "sensitive data," "data sharing," etc.
- explore the construct of the data privacy calculus — trust, control and benefits — to understand your own data-sharing practices
- draft a statement on learner consent for your local institutional or course context

Ethical issues: a brief overview

Ethical implications around the collection, analysis, and uses of learner data should take account of the potentially conflicting interests of a range of stakeholders, such as learners and their institutions. Views on the benefits, risks and potential for harm from the collection, analysis and use of learner data will depend on the interests and perceptions of the particular stakeholder. In this section, we briefly introduce a number of practical considerations. First, have a look at the material in the box, which we will revisit later in the unit.

Look at the following statements and then decide which meet the following question:

i. Consent to use learner data must be obtained at the point of registration.
ii. Once an institution can predict a likely learner outcome, it must act upon that insight.
iii. Learning analytics can only be used by trained analysts or statisticians.
iv. Once a learner registers, all of their data (e.g., provided at registration or later click data) becomes the property of the institution.

Which of these do you think is true?

1. All of them
2. (a) and (d)
3. (a), (b) and (d)
4. Only (a)
5. Only (d)
6. None of them

There are many ways to consider and categorise ethical issues. The following headings explore some of the issues you may wish to consider. (There are others!)

What data are collected, for what purpose, and who has access to it?

This concerns categories of data and intended use. That is, data should only be collected and analysed if they are relevant and required for the proposed learning analytics activity. Speculative data collection (harvested in case it should become of interest in the future) should be avoided. Legitimate use or purpose is typically defined as being specified, explicit and sanctioned. Such uses or purposes should ideally be set out in advance in such a way that all stakeholders understand them. Data collected for legitimate purposes should only be used for those purposes and not be further processed for other purposes; in an educational context, admissible processing would include what is needed for archiving or for research. In the absence of a clear policy, it is relatively easy for an institution to overlook the need to restrict data collection, analysis and use to a contained set of purposes. One example might be a higher education institution indicating (formally or informally) that learner (demographic and behavioural) data are collected and analysed specifically for learning design purposes, but then going on to share subsets of that data with third-party marketing companies. Most commonly though, there is the issue of finding useful things in a data set that push the use into an area not previously covered or agreed.

Some national and federal/state legislation prohibits or constrains the collection of personal or sensitive data, and you should ensure that you are aware of what data types fall into this category in your own context. For example, in Canada, "sensitive data" are not specifically defined under national data protection legislation or provincial data protection statutes, although data can be classified as sensitive, depending on the context. In the European Union, however, sensitive data are defined as: data consisting of racial or ethnic origin, political opinions, religious or philosophical beliefs, trade union membership, genetic data, biometric data, data concerning health or data concerning a person's sex life or sexual orientation.

Reflective action: become aware of relevant legislation

Depending on where your institution is based (and whether it has multiple international sites or third-party suppliers that transfer learner data between them), research the relevant national and federal/state legislation to identify sensitive and/or personal data categories that may not be collected and/or analysed. Assess whether there is any definition of legitimate use formally set out, as well as limits on how data may be shared between the institution and others. In addition, research whether there are any institutional policies in place that limit the collection and sharing of any data categories. Make a note of these.

Similarly, you should consider who has access to data and why they would need it. You might consider that “raw” data — e.g., taken directly from a learner’s record or from their online log files — should be restricted to those with a need to know. For example, support or administrative staff may need to understand a learner’s personal situation in order to provide relevant personalised advice and support. It is less likely that a course designer would need similar access to that raw data, although they might want to see processed, aggregated data for a learner cohort or class to better understand how to redesign an assessment strategy.

Reflective action: who and why

If you’ve enrolled in this course, there’s a good probability that you like data — understanding the patterns you might find, gaining insight from analyses, etc. However, although it is tempting to collect data and then see what you can extract from it, you should aim to reverse that thinking. Consider what you (or your institution) is seeking to achieve. As you did in Unit 3, record the purpose(s) of learning analytics in your context.

Note down all of the stakeholder groups that might have some involvement in learning analytics. Now think about the data categories available to you (excluding those considered off limits, as discussed above), and note down which categories of data are relevant and likely available to each stakeholder group. For example, you will almost certainly consider instructors or teachers to be a relevant group; what data do they have access to? What data might be considered out of bounds (irrelevant or sensitive)? We will revisit this activity in somewhat more detail in the next unit.

Transparency and consent

Transparency focuses on the extent to which an institution is open and clear about the purposes of learning analytics. This should include making those purposes accessible; there is little point in having a policy (see Unit 9) if learners and other stakeholders are unaware of it. Although setting out the purposes of learning analytics may feel obvious, it is fair to say that its uses are often made as much for the benefit of the institution (maximising throughput) as the individual learner (providing the "best" outcome, whether that relates to improving course scores or allowing free study choices).

Transparency also includes making clear what data are collected (and what are not), and any assumptions made about those data (where they may be incomplete or acting as a proxy for another measure, for instance). Raising awareness may itself introduce challenges, since many are often unaware of the extent to which data are now routinely collected. Of course, it is not always possible to be completely transparent — models built around regression approaches, for example, can be difficult to understand and interrogate. It is not always clear why one learner is identified as potentially more vulnerable than another. Nor is it always in the best interests of learners to communicate a predicted poor outcome (unless this is appropriately balanced with the provision of additional support or alternative study paths).

Transparency is necessary for meaningful consent. Often, consent to collect learner data for learning analytics (as opposed to collection for normal administrative uses, say) is sought at the point of registration, if at all. At this point, many learners will be largely unaware of learning analytics and how it may be used to provide ongoing insight. Consent at this point is certainly most convenient for the institution, but arguably less meaningful for the learner (unless there is an opportunity to later withdraw it). Various frameworks and policies propose different approaches here, often influenced by existing legislation in the national context. One broad approach might be to differentiate between initial broad consent for the collection of data and specific consent when data are used to intervene in the choices learners have, when adapting their learning experience, or when access to resources is suggested. Although there are practical difficulties in doing so, an expectation that users should consent to uses of personal data unknown at the point of registration is perhaps an unreasonable and unethical one.

Reflective action: review your terms and conditions

Find the terms and conditions statement for your own institution (or if you can't easily find it, take a look at the terms of use for US MOOC provider edX, [28] and its embedded privacy policy[29]). How accessible do you find the terminology? Does it describe or define what type of data will be collected, how that data might be used and who the data could be shared with? Are there alternatives if a user objects?

[28] https://www.edx.org/edx-terms-service

[29] https://www.edx.org/edx-privacy-policy

Data storage

Data storage and security are usually determined by existing data protection policies (check yours to see whether you think it is clear and detailed enough for most learners to understand). Additional issues for consideration include data sharing between the institution and other third parties — usually those offering a service, such as marketing. Some institutions effectively outsource their learning analytics activities, either wholly or in part, by using proprietary software that may capture and store learner data. It is sensible to ensure that your institutional approach leaves learner (and other) data secure and not open to unintended consequences.

Data ownership

In a learning analytics context, the institutional presumption is often that data collected are owned by the institution. Some, though, see data as an extension of the learner's identity; for example, "learner data is not something separate from learners' identities, their histories, their beings. . . . [it] is therefore not something a learner owns but rather is" (Prinsloo, 2017). And although there is often clear existing legal protection relating to personal data, the lack of clarity around who owns the data muddies principles of meaningful consent. Another perspective then might be to assume that the institution has temporary stewardship of its learner data. Subject to legislation and policy, the institution may store data sets under certain conditions and for specified periods, but within a higher education context, the issue of ownership could remain open. In this situation, the institution would be able to collect and analyse learner data, and apply insights from learning analytics, but would be constrained from further exploitation of that data.

Interpretation issues

If analytics is predictive (i.e., not built only on known events, but involving statistical calculations), data sets should be complete and sufficient to ensure any calculations are robust. Further, the models used to analyse, interpret and communicate learning analytics to stakeholders (learners, teachers, support staff, advisers, etc.) should be free from algorithmic bias, transparent where possible, and clearly understood by the end users. This is not to say that everyone needs a postgraduate qualification in statistics, but we should be aware of the limitations of the outputs and understand that prediction is not equivalent to fact. Analytics outputs can be misinterpreted or misunderstood if the end user lacks the skills to extract clear meanings and causal relationships.

Some robustness may be lost when, for example, the thing that the user wants to get a feel for or measure is hard to quantify, so an available proxy is used instead. A classic example here is the measure for learner engagement — often extracted from time spent online as an achievable but arguably flawed measure.

The obligation to act

Should access to knowing, and understanding more about how our learners learn and how they are progressing, equate to a moral obligation to act? For example, having observed learners not submitting summative assignments, or having calculated the probabilities of course completion, is the institution obliged to act on what it has identified? Often, resources (usually staff time) are constrained, and it is not always easy to reach all learners identified as likely to benefit from additional support. In these cases, the institution might consider a type of “triage” policy by focusing available time and support on the group identified as, for example, most potentially vulnerable; or on learners in high-population courses; or on learners with particular characteristics (for example, those with known disabilities). Realistically, a line has to be drawn. However, the decision-making process should be transparent and align with other known strategies.

Data privacy calculus

Privacy calculus theory suggests that individuals may be willing to share some or all of their data after evaluating the potential benefits and risks. Research by the Pew Research Center [30] in the US indicates a range of views around the acceptability of organisations sharing personal data — most often depending on the organisation and the perceived purpose. For example, their 2019 report states that "49% [of Americans] say it is acceptable for the government to collect data about all Americans to assess who might be a potential terrorist threat," whereas "about four-in-ten are concerned *a lot* about what personal information social media sites . . . might know about them." Research carried out in 2018 at The Open University, in the UK, looked at how learners viewed their own privacy calculus whilst also measuring what they *actually* did in practice. This research found that very few learners ever read the terms and conditions of websites visited, but that "93% of respondents indicated that it is 'quite' or 'very important' to be in control of who gets access to information shared online" (Slade et al., 2019). In the context of education and learning analytics, there was a higher level of trust (as might be hoped), with around three-quarters of learners stating that they trusted the university with their data. In reality, of course, user preferences can be irrelevant, since the options are most often limited to sharing data or being denied access to the service.

[30] https://www.pewresearch.org/internet/2019/11/15/americans-and-privacy-concerned-confused-and-feeling-lack-of-control-over-their-personal-information/

For more insight into the privacy calculus, watch this short video.

Watch Video: https://youtu.be/k0PW_5JlF88

Video attribution: "Unit 7: The Privacy Calculus" by Commonwealth of Learning is available under CC BY-SA 4.0

Reflective action: how do you share?

Think about the websites that you visit most often. Have you read their terms and conditions? (Be honest!) Do you know what data are collected by the site and how these might be shared? Would that matter to you? Most people have some concerns or feel some unease at the thought of their data being shared with other parties but often feel that access to the service is sufficient payment (or feel that the risk is low enough not to worry).

Reflection action: learner consent in your context

Having reached this stage in the course, you will have a better understanding of how learning analytics can be applied in educational settings, what data may be collected, and a range of ethical issues. Consider the issue of consent. If you were to take a position on consent in your own institutional (or course) context, what would you include?

Summary and conclusion

This unit has given a quick overview of the complex area of ethical concerns over uses of (learner) data. None is necessarily more important than any other, and views on each will vary, depending on the institution and the stakeholder. Unit 9 considers how some of these issues can be formalised via the creation of a learning analytics policy.

Suggested readings

There is a growing body of research summarising these issues in learning analytics. The following are some useful readings.

Learning analytics: ethical issues and dilemmas: http://oro.open.ac.uk/36594/2/ECE12B6B.pdf

GDPR and Learning Analytics: frequently asked questions: https://analytics.jiscinvolve.org/wp/2018/06/01/gdpr-and-learning-analytics-frequently-asked-questions/

Ethical challenges for learning analytics: https://epress.lib.uts.edu.au/journals/index.php/JLA/article/view/6587

Practical ethics for building learning analytics: https://bera-journals.onlinelibrary.wiley.com/doi/abs/10.1111/bjet.12868

Check your progress

a. Having read a summary of potential ethical concerns, let's revisit the following statements and again make a decision:

   i. Consent must be obtained at the point of registration.
   ii. Once an institution can predict a likely learner outcome, it must act upon that insight.
   iii. Learning analytics can only be used by trained analysts or statisticians.
   iv. Once a learner registers, all of their data (provided at registration or click data) becomes the property of the institution.

   Which of these is always true?

   1. All of them
   2. (a) and (d)
   3. (a), (b) and (d)
   4. Only (a)
   5. Only (d)
   6. None of them

b. Learning analytics can predict whether a learner will pass a course.

   i. True
   ii. False

c. What is the data privacy calculus?

   i. a weighing up of the pros and cons of using learning analytics
   ii. a weighing up of the pros and cons of sharing data in order to access a service
   iii. a devilishly tricky math calculation

d. What is needed for meaningful consent? (choose all that apply)

   i. a formal policy statement
   ii. transparency about uses of learner data in learning analytics
   iii. a robust approach to using algorithms
   iv. an ability to change consent positions at later stages

UNIT 8

Implementing Learning Analytics

Introduction

Welcome to Unit 8. While student data have long been collected and used, there is additional work associated with the widespread implementation of learning analytics. There is a need to understand who the stakeholders are, to identify the systems and processes needed to implement learning analytics, and to then design and operationalise responsive and integrated systems and processes. Together in this unit, we explore the dimensions that will put you one step further into applying learning analytics for your course or institution!

Learning outcomes

After you have worked through this unit, you should be able to:

- conduct a SWOT analysis of your local institution/course context for the implementation of learning analytics
- explore various open applications for implementing learning analytics in your own (course) context
- compile a list of stakeholders that you would invite to an information session on the potential to implement learning analytics, or to further improve and increase the impact of learning analytics

Learning analytics stakeholders

In Unit 7, you were asked to think of who the stakeholders in learning analytics might be. You may have thought of who contributes to the model of learning analytics in your class, or within the wider institution. We'd guess that you thought of two key players at least: learners and teachers. But there are many others who routinely engage with learning analytics. In fact, stakeholders will differ, depending on the context, e.g., higher education and primary schools. Yet common key players take on similar roles.

According to two studies (Drachsler & Greller, 2012; Leitner et al., 2017), four common stakeholders appear repeatedly when implementing learning analytics, regardless of the context: learners, teachers/tutors, researchers/administrative staff, and the institution itself. Other stakeholders that might also be involved include course designers, data keepers, governments, parents, senior leaders, teaching assistants, quality assurance, information technology staff, marketing staff, and privacy and ethics governors.

One helpful approach to identifying your key stakeholders is to follow the questions within the SHEILA framework.[31] We will refer to the section covering stakeholders next and dig deeper into SHEILA in the following section.

[30] https://sheilaproject.eu/wp-content/uploads/2018/08/SHEILA-framework_Version-2.pdf

Case study on identifying stakeholders in learning analytics

Laila is a learning analytics expert who has started to work at a local school. The senior management team asked Laila to start planning for implementing learning analytics at their school, since learning and teaching processes have been rapidly moved online. The school's major objectives are to provide performance metrics to parents and to intervene when learners appear to fall within a danger zone of failing in classes. Laila wants to identify the key people she needs to talk to so that the implementation of learning analytics at the school is successful and achieves the proposed objectives. She uses the SHEILA framework to identify the personnel as follows:

Identify primary users of learning analytics

Laila selects learners and teachers as the main stakeholders for her approach.

Identify senior management

There are many levels of management at the school, including the head teacher, deputy principal, assistant principals (curriculum for learning, standards for learning), school managers, and a digital learning advisor. Laila plans to talk to each to see where their decision-making responsibilities and insight will support her implementation.

Identify professional team

She selects the IT support, student welfare and support staff, and a key member of the school office team to support her here.

Select academic team

Since the context of the learning analytics in the school is computer-based, she involves the rapidly formed digital-learning group and the department leads for each curriculum area.

Identify external parties

Laila decides that the parents of the learners are fundamental, given that they have the legal guardianship of the learners involved and will be recipients of some of the metrics.

Identify required expertise

Laila singles out other expertise needed. She wants to involve didactics expertise, IT expertise, learning analytics expertise and statistics expertise to help the professional team in carrying out the implementation.

After spending four months working on this project, Laila was able to identify the major parties needed to achieve the school's goals of implementing learning analytics. Selecting the right stakeholders was a great start in helping Laila and the school support learners and parents.

SHEILA framework

SHEILA is a project that supports European higher education institutions in managing and using their digital learner data. By conducting a broad range of in-depth interviews and surveys with learning analytics stakeholders, teachers and learners, the project developed a framework to assist higher education institutions to develop, implement and evaluate their uses of learning analytics. The SHEILA framework provides a sound foundation on which you can base the various stages of your own learning analytics approach.

To support you in this, we summarise the framework and highlight the major components.[32] figure "SHEILA framework dimensions and components" illustrates the six dimensions of the SHEILA framework. Each dimension also includes a relevant action, challenge and prompts for use in your future policy development.

SHEILA framework dimensions and components

[30] https://sheilaproject.eu/sheila-framework/create-your-framework/

Dimension 1 – Map political context:

The first dimension asks you to identify the internal and external drivers for learning analytics:

- **Action**: Think of the political context and identify opportunities.
- **Challenges**: What are the challenges that may exist in the political context?
- **Policy**: Think about reasons for adopting learning analytics and how your institutional objectives line up with those reasons.

Dimension 2 – Identify stakeholders:

In the second dimension, you need to identify key stakeholders. We have touched on this already in Unit 7 and revisited it briefly in section "Learning analytics stakeholders" of this unit.

- **Action**: Think of the key personnel to implement the learning analytics model.
- **Challenges**: What are the challenges of involving some of the stakeholders? Are there any risks if you want to involve learners?
- **Policy**: Think about how you will get consent. (Hint: re-read Unit 7).

Dimension 3 – Desired behaviour changes:

Identify the changes to your institution, course or stakeholders that might arise when implementing learning analytics.

- **Action:** Think of the changes that could occur when learning analytics is implemented.
- **Challenges:** Think about ethics and privacy, infrastructure, capabilities, management, etc.
- **Policy:** How might you communicate with relevant actors about the changes?

Dimension 4 – Develop engagement strategy:

- **Action:** What would you need to carry out the changes in dimension 3? Think of who to appoint. Does your project need funding? Think about resources for that!
- **Challenges:** What are the challenges that will hinder your development of the engagement strategy?
- **Policy:** What type(s) of data do you collect, analyse and report? (Hint: see Unit 4.)

Dimension 5 – Analysis of the internal capacity to effect change:

How ready is your institution to change based on dimensions 3 and 4?

- **Action:** Think about evaluating your institution/course infrastructure to handle any changes. Would this need training? Would there be any cultural issues to consider?
- **Challenges:** Think about human capital and infrastructure. What are the challenges that might slow or prevent the planned changes?
- **Policy:** What changes might be needed to your institutional policies?

Dimension 6 – Monitoring and learning frameworks:

Establish indicators and measures for a successful implementation.

- **Action:** List indicators and measurable metrics.
- **Challenges:** Fail to recognise and address limitations of data and analytics models.
- **Policy:** Identify the limitations of learning analytics in your institution — i.e., what can be done and what not?

Learning analytics SWOT matrix

At this stage, we have covered a lot. You have learnt about data-driven approaches in education; the evidence of data; the difference between learning analytics and other derivations; uses of learning analytics; tools; frameworks; and ethical and privacy concerns. Since this unit focuses on implementing learning analytics, we would like you to carry out a short activity to help you prepare. On your own device (or even using a pen and paper!), identify the strengths, weaknesses, opportunities and threats related to your course or institution when implementing learning analytics. Use the matrix below to guide your thoughts.

Strengths **S** 1. Improve my course design 2. 3. 4.	***Weaknesses*** **W** 1. Lack of real-time data communication 2. 3. 4.
Opportunities **O** 1. Predict learners who might fail 2. 3. 4.	***Threats*** **T** 1. Exposing my learners' private data 2. 3. 4.

Learning analytics implementation in my course and/or institution: SWOT matrix

Case studies

Have you ever wondered about how learning analytics work with educational data? Do you want to try some off-the-shelf tools? Then welcome to this part of the course! Unit 6 has already highlighted a number of available tools, but here we provide another that you can quickly experiment with.

EdX Logfile Analysis Tool (ELAT)

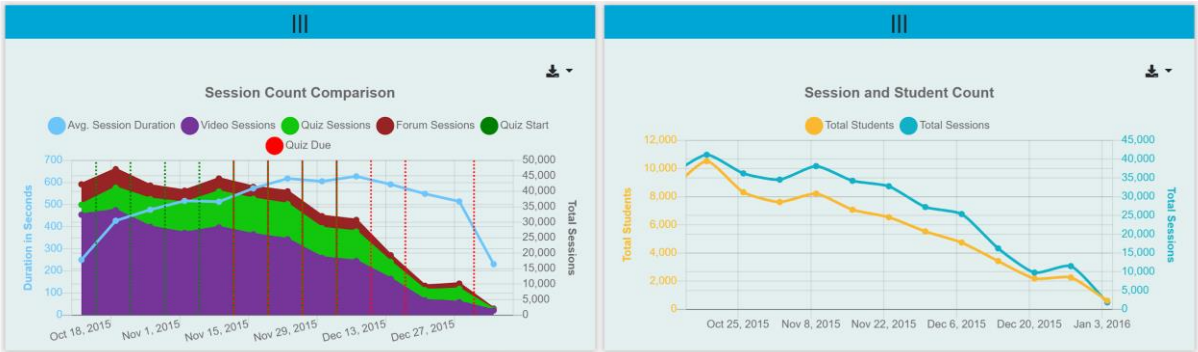
As you will now know, learning analytics is concerned with small and big data. Sources of data vary in size and level. We would like to provide you with an example from the world of MOOCs learning platforms. As their name indicates, MOOCs can have lots of students, content and data!

Activity test a web-based tool

The Technical University of Delft, in the Netherlands, has developed a tool (ELAT[33]) that analyses data from the edX[34] MOOC platform (see Figure "ELAT snapshots"). You do not need advanced skills to use ELAT, nor do you need data! This tool is browser based and provides a sample data set for you to play with. We are delighted to share it with you. Access the tool via https://mvallet91.github.io/ELAT-Workbench/, where you are offered a range of descriptive statistics and advanced functionalities, including machine learning and static and interactive visualisations. Check out the tool and navigate around. Use the functions of ELAT, and read the quick guide in case you need further help.

[33] https://mvallet91.github.io/ELAT-Workbench/

[34] http://edx.org



ELAT snapshots

OnTask

In this section, we share another principal case study of learning analytics implementation. OnTask[35] is an open-source tool developed by several Australian universities and funded by the Office of Learning and Teaching of the Australian Government. OnTask aims to improve the learning experience by providing learners with timely, personalised and actionable feedback throughout their course. The software can be downloaded and used on your local machine. There is also a discussion forum available for inquiries if you need help. See Figure "OnTask screenshot" for a screenshot of the tool.

So how can you use OnTask? Here's one example of how it might help:

Scenario:

You are a teacher who wants to send customised messages by e-mail to your learners based on specific conditions. You create three customised messages for your learners who

a) have not introduced themselves to their peers,

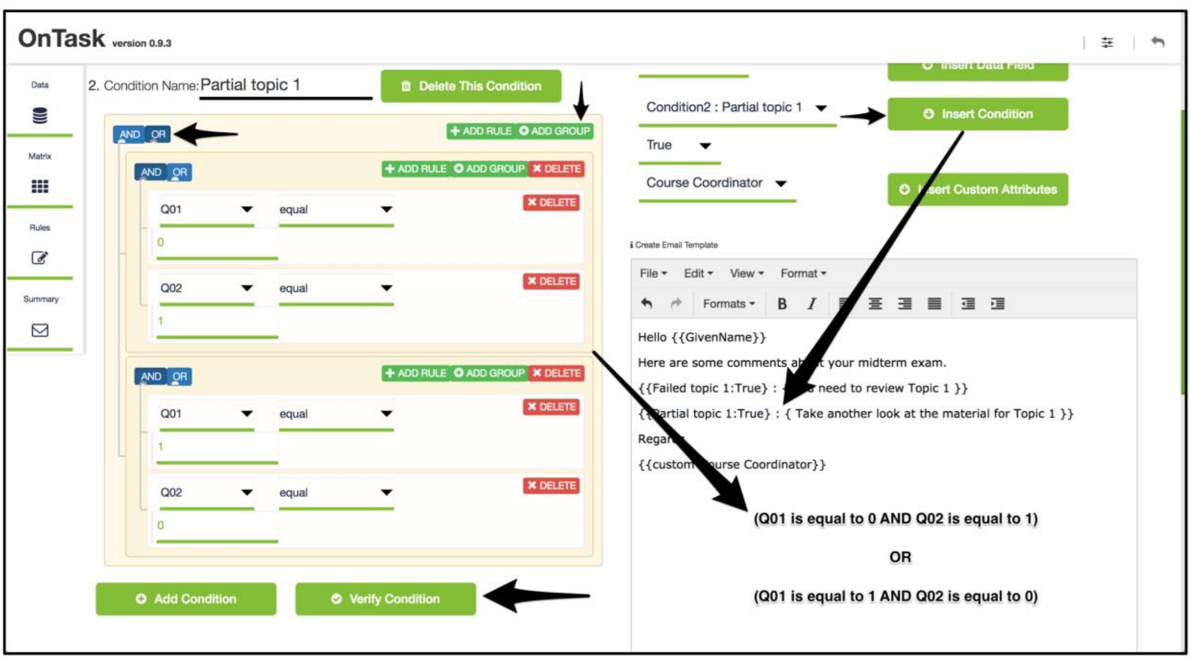
b) have not submitted their assignments on time or

c) have not watched at least 75% of the course videos.

OnTask provides "if-this-then-that rules" that you can use for those cases. In very simple terms, you have the following options here:

[35] https://www.ontasklearning.org/

- For case (a), you program the tool to send "Dear *student x*. You'll find it helpful to discuss course issues in the forum. As soon as you can, try to make sure you introduce yourself to your peers."
- For case (b), you program the tool to send a message stating: "Well done for getting your assignment in, although you missed the course deadline. Please aim to submit the next assignment of *course x* on time. If you are having problems keeping up, please do contact me."
- For case (c), you will want to use three conditions — let's say: if the learner fails to watch any videos, or views up to 50%, or watches at least 75%. You teach the tool to send a tailored message: If a learner never watches the videos, then "Dear *student x*. In order for you to answer the self-assessment questions, it is really important that you watch the *course x* videos"; if *student x* watches up to 50%, then, "*Student x*, keep up the good work! There are still more videos for you to watch." The final condition could trigger, "*Student x*, I've noticed you have watched most of the course videos. Great work — keep going!"

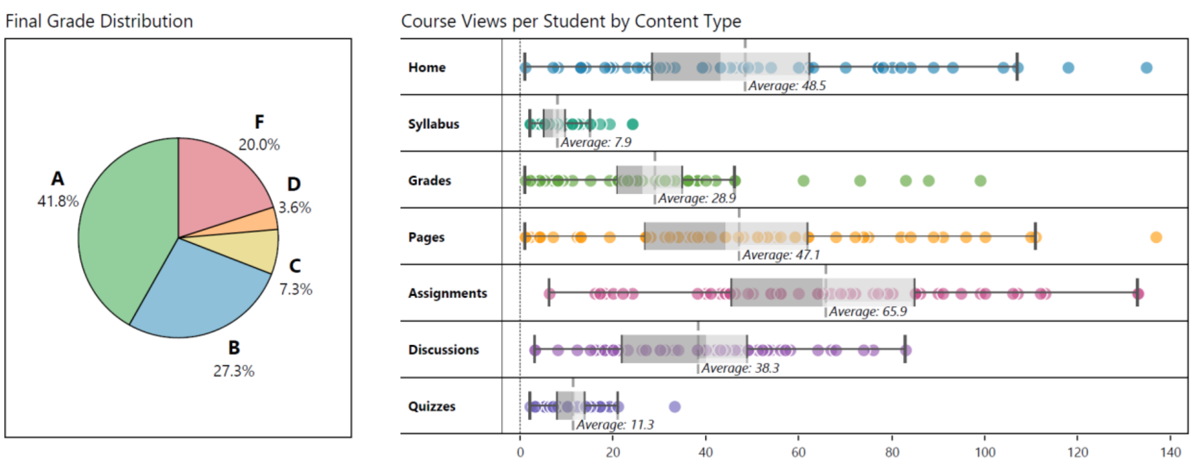


OnTask screenshot[36]

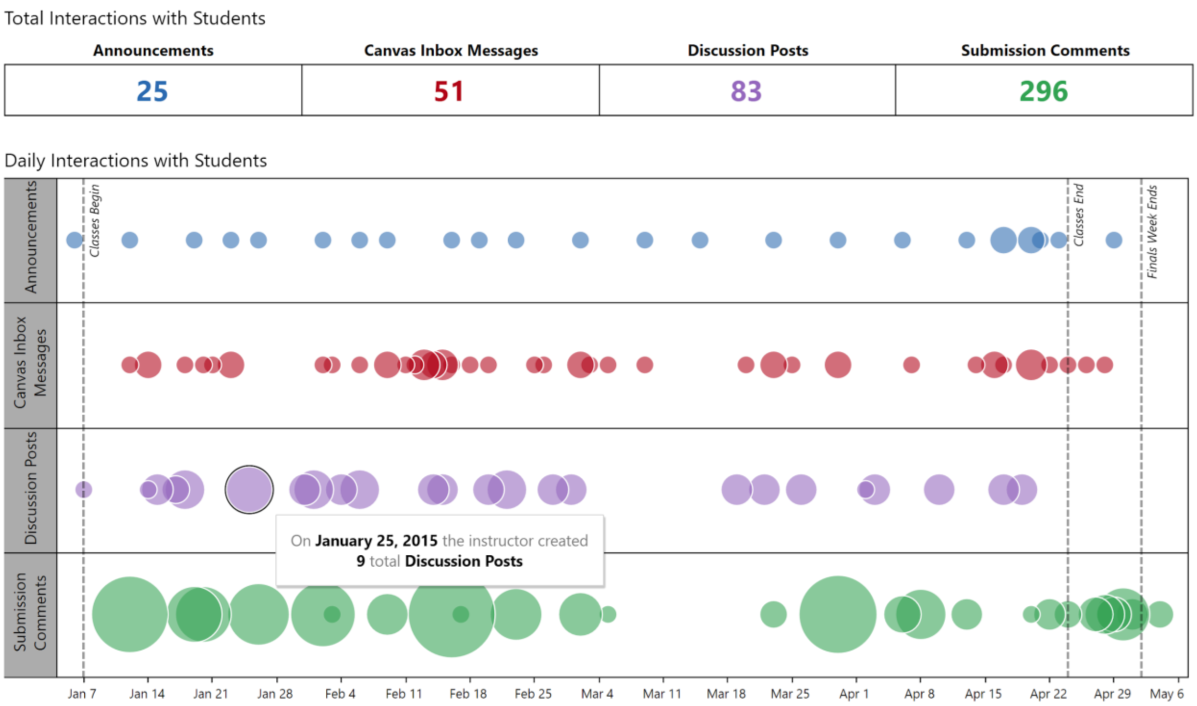
[36] https://www.ontasklearning.org/wp-content/uploads/OnTask-Brief-Description-v1.pdf

Analytics for teaching

The path to good learning starts with good teaching. In this example, we highlight how learning analytics can be used for documenting teaching and learning through dashboard designs. An initiative from Utah State University [37] shows designs of visualisations and analytics that could be used to review and improve learning, teaching and course designs (Figure "Two dashboard views of analytics from Utah State University" provides examples of two dashboard views).



[37] https://cidi.usu.edu/Analytics



Two dashboard views of analytics from Utah State University.

The top figure shows the relationship between learners' final grade and their course views. The bottom figure shows teacher interactions with the content and learners over time.[5]

A teacher engaging with the first dashboard view above may use the links between student engagement with different aspects of the course and the final grade distribution. They might then reflect and opt to restructure certain components, by editing content, say, or by leaving messages on less-viewed pages.

The second dashboard view provides the teacher with a timeline of his/her interactions with the learners in the LMS, including discussion forums, comments and announcements. They can quickly see where they are spending most time at various stages and where further support or signposting might be needed.

For another example and more insights into learning analytics implementation, watch this short video.

Watch Video: https://youtu.be/KGe82vXjsIQ

Video attribution: "Unit 8: Learning Analytics in Action" by Commonwealth of Learning is available under CC BY-SA 4.0.

Open data sets for implementing learning analytics

Do you want to try learning analytics, but you lack data? We know that getting your hands on robust data sets is not easy. However, there is a movement to share data, which is similar to the open educational resources (OER) and open research movements. The table below lists some sources of open data sets that you may wish to explore when playing with learning analytics approaches and tools:

Data set link	Description
https://research.moodle.org/view/subjects/	A collection of publications of Moodle-related research
https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/26147	De-identified data from the first year (Academic Year 2013: Fall 2012, Spring 2013, and Summer 2013) of HarvardX courses on edX
https://dataverse.harvard.edu/	Search a wide range of data sets available from Harvard
https://datasetsearch.research.google.com/	Google search engine that indexes several data sets on the Web
http://www3.dsi.uminho.pt/pcortez/Downloads.html	Data sets including educational ones provided by Prof. Paulo Cortez of University of Minho, Portugal

Open data sets for exploring learning analytics

Summary and conclusion

Toward the end of this unit, we covered important aspects of implementing learning analytics. We discussed the advantages of considering stakeholders and shared a case study to illustrate how this might be done. The SHEILA framework was introduced, providing you with a clear and comprehensive framework. We also provided three examples illustrating tools and approaches implemented at educational institutions in different countries (the Netherlands, Australia and the US). Finally, we have pointed you toward a set of free and open educational data resources that you may use to try learning analytics! Thank you, and good luck with Unit 9: Developing Policy for Learning Analytics. You are almost there!

Check your progress

a. Researchers are identified as important stakeholders for learning analytics when their roles involve evaluating a learning analytics system.

   i. True
   ii. False

b. To improve the learning experience at one institution, the learning analytics expert decides to depend largely on interviewing learners to design an effective learning analytics model.

   i. True
   ii. False

c. Data surveillance is one of many challenges facing learning analytics implementations.

   i. True
   ii. False

d. According to the SHEILA framework, identifying stakeholders is

   i. the first step you should do when implementing learning analytics.
   ii. where you can identify areas that might be supported by learning analytics.
   iii. where you explore who will be responsible for data controlling.
   iv. where you can seek funding for your learning analytics project.

e. Despite data sets being a key corporate asset, they are generally not given the importance they deserve, in terms of making them easy to find and accessible.

   i. True
   ii. False

f. Which of the following is an example of the third dimension of the SHEILA framework (desired behaviour changes)?

   i. Consult relevant policies and codes of practice to implement learning analytics.
   ii. Choose analytical models and define measurements of success for your learning analytics approach.
   iii. Identify drivers for learning analytics and the objectives behind its implementation.
   iv. Adjust the teaching method to fit the implementation of learning analytics.

g. Involving IT staff in your learning analytics implementation is an example of bringing in the:

   i. professional team stakeholder.
   ii. academic team stakeholder.
   iii. external party stakeholder.
   iv. professional and external stakeholders.

h. There are several available frameworks to support the regulation of learning analytics within a policy context.

   i. True
   ii. False

i. There are many learning analytics applications that only function on a single digital/computer-based learning platform.

   i. True
   ii. False

j. You are asked to map the political context of an existing learning analytics system that seems to be suffering from data surveillance issues. Which of the following would you prioritise?

   i. Identify the stakeholders that can help you mitigate the issue.
   ii. Contact the IT team and stop the system completely.
   iii. Study the reasons and identify the drivers and the institution's strategy to mitigate the problem.
   iv. All of the above.

Section 9

Developing Policy for Learning Analytics

Introduction

Having now studied several units of this course, you'll be more aware of the extent to which institutions collect, analyse and use student data. For many institutions, formal guidance or policy on how and why this is done may not yet exist, particularly if uses of learning analytics are relatively new or have not yet begun. For others, existing policies relating and referring to potential uses of student data may need fresh scrutiny to ensure their continued relevance and completeness. This unit sets out some of the steps involved in creating a policy framework, advises on key principles and provides examples of existing educational policy that might provide a useful starting point. Having studied the unit, you should be able to prepare a preliminary outline of the steps involved and what a policy might look like in your own context.

Learning outcomes

After you have worked through this unit, you should be able to:

- explain the key features and purpose of a learning analytics policy
- compare and contrast learning analytics policies at the national and institutional levels
- use the template provided to draft the outline of a context-specific learning analytics policy

Understanding the purpose of a learning analytics policy

It's fair to say that many educational institutions have a whole pile of formal policies tucked away on their website (or perhaps not even open for scrutiny) that very few ever read. Yet the purpose of having a visible policy is multifaceted.

Fundamentally, a policy should make clear to all stakeholders what the institution's position is on the issues. It should explain the background and the need, establish any boundaries (what is included and what is not) and justify assumptions. Any new policy should support and not conflict with other existing policies (or require that those be revised to remove that conflict). Unlike guidance or a framework, it will do more than suggest, but seeks to clarify acceptable practice and make formal the "rules" around a particular issue.

Learning analytics is a relatively new practice for many educational institutions. Student data have been collected for many years, often for reporting or operational purposes, but the range of data now collected for learning analytics, and the reasons for its collection, warrant a fresh look at whether additional policy is required.

A hidden benefit of policy creation is that it forces key stakeholders to fully consider the range of issues — the context; the boundaries; the purpose of the process/activity; who it is important to; who needs to be involved in the decision making; and what needs to be covered. Policy creation is no easy exercise; it requires rigour, research and resilience!

Steps toward a learning analytics policy

The steps needed to create a new policy will depend, in part, on the size and complexity of the organisation, current and planned future uses of learner data, existing policies, and relevant (inter)national or federal law. The following steps may not always apply, but it is recommended that they be at least considered in order to produce a robust outcome relevant to your own context. Ideally, policy should be set out and agreed *before* the introduction of learning analytics in your institution, as it should guide the application of learning analytics in your contexts. Realistically, as more institutions adopt some form of learning analytics, policy creation is likely to be a retrospective process. *In such cases, extra care should be taken to ensure that ongoing learning analytics are driven by the final policy rather than the other way around*.

1. Form your team

Start with a small team (preferably five or fewer). Ensure that you have a project lead and a project manager. You will also need members who understand some or all of the following: current or planned uses of learning analytics; data protection issues; how policy is created within your organisation. Be prepared to meet regularly — e.g., monthly — and be clear about what is expected from each member. Keep a well-defined record of all discussions and agreed objectives. Maintain a shared space for collecting and sharing relevant resources. Agree on a schedule and a meeting frequency.

2. Agree on and set the purpose

This is perhaps the most obvious and yet potentially the most overlooked stage of creating any policy. Before the organisation begins to draft any new policy, it must first understand what it is trying to achieve.

In the case of implementing a learning analytics policy, the purpose might be very specific — for example, to improve learner completion or pass rates at a course level. Or perhaps a more general aim would be to better understand how learning design has impacts on learner success. Once discussions have begun, it will become clear that boundaries need to be set. For example, are there any activities or groups of stakeholders that must be included or would be excluded from the policy? In the case of learning analytics, you may wish to consider whether teacher data are included, for example, or whether all data collected by the institution will always be used. An example of the latter might be the following statement:

All data captured as a result of the institution's interaction with the learner has the potential to provide evidence for learning analytics. Data will only be used for learning analytics where there is likely to be an expected benefit to learners' study outcomes.

It is perhaps better to develop a purpose that covers current (or planned) learning analytics activities than to try to cover all future eventualities. Learning analytics is a fast-moving field. Start simple. If your organisation focuses on, for example, providing interventions for learners not meeting certain deadlines or tasks, it will be simpler to define your aims and scope than if you assume that it will at some point move into machine learning.

3. Research

It is essential to understand the organisational context in which the policy will exist. Time should be spent investigating existing policy, and understanding how uses of learning analytics are likely to have an impact on your organisation's aims and the needs of key stakeholders going forward. Key related policies here (depending on your particular institution) are most likely to be related to data protection and data retention.

You may consider whether the uses of a range of technologies also comes into play here — what technologies (e.g., apps used on mobile devices, personal computers, websites, etc.) are used to deliver learning in your setting. How are learner data captured and shared, and with whom?

At this point, it would be useful to review existing policies and best practice for uses of learning analytics in other comparable learning institutions. Fundamentally, some of your consideration should be directed to understanding how existing and planned legislation may constrain your planned approach, as well as specific data sets that may be out of bounds for certain analytics activities (as discussed in Unit 7). In many countries, you will certainly need to understand what is meant by "personal information" and/or "sensitive information," as these categories are often more constrained in terms of what can be collected and used within learning analytics. An example of personal information that may not be collected in some US states is a student's social security number. In the European Union, personal information includes any information that relates to an identified person or one who could be identified, directly or indirectly, based on the information. Whether it may then be used for learning analytics purposes could then depend on whether it is further classified as Special Category Data (e.g., religious or philosophical beliefs).

Here are examples of data protection legislation in a range of geographic contexts:

- The Personal Information Protection and Electronic Documents Act (PIPEDA) governs how private-sector organisations collect, use and disclose personal information. However, for public sector bodies, such as schools, provincial and territorial privacy laws apply.[38]

[38] https://www.priv.gc.ca/en/about-the-opc/what-we-do/provincial-and-territorial-collaboration/provincial-and-territorial-privacy-laws-and-oversight/

- The Family Educational Rights and Privacy Act of 1974 (FERPA[39]) addresses the privacy of student education records in the United States. Note that state law can also apply in the US — for example, the California Consumer Privacy Act 2018 (CCPA[40]).
- The General Data Protection Regulation (GDPR[41]) in the European Union.
- In Australia, the Privacy Act (1988) [42] regulates how government agencies (amongst others) handle personal information, although again, other state and territorial legislation may also apply.
- You should also consider whether any aspects of organisational practice cross (or are likely to cross) national or state legislative boundaries. For example, are there any activities that involve data transfer outside of your country? How might this affect your policy statement?

4. Identify key elements and draft the structure

Based on your understanding of the purposes of using learning analytics and of the relevant policies and legislation, you should now be able to establish an outline for your new policy. A suggested outline might contain sections such as:

- Introduction/background
- Definitions (of terms used, etc.)
- Scope/boundaries – what learning analytics activities and data types are included in the policy, and which are out of scope; who can access student data and any resulting analytics within the institution

[39] FERPA: https://studentprivacy.ed.gov/node/548/

[40] CCPA: https://oag.ca.gov/privacy/ccpa

[41] GDPR: https://gdprinfo.eu/

42] The Privacy Act: https://www.oaic.gov.au/privacy/the-privacy-act/

- Relevant institutional policy and legislation
- The principles underlying the policy
- Implementation strategy and timeline
- Review process and policy owner(s)

5. Stakeholder input and ongoing review

Much of the time taken to develop a meaningful policy is invested at this stage. If your new policy is to get genuine buy-in, you need to engage with all representative and relevant stakeholders. Consider: who does learning analytics impact, and who is driving the agenda to adopt learning analytics within your organisation? This is really an iterative process, but you should first meet with different stakeholders to capture their views on issues such as acceptable uses of learning analytics and data types, and then revisit those groups once you have developed sufficient content. This is not an all-inclusive list, and remember, you have considered relevant stakeholders for learning analytics in Unit 8, but your stakeholders might typically include:

- **Learners**. Is there a recognised panel or body in place? If not, you might consider a consultative forum or survey to gather input. Don't be tempted to skip this group; they are perhaps the most important of all stakeholders. Be aware that raising the issues around uses of learning analytics may create unease for some, so be prepared and be clear on the purposes, boundaries and data categories.
- **Course leaders/course teams**. How might they use learning analytics?
- **Information technology staff**. Data storage issues, etc.
- **Management**. Depending on the size and structure of your organisation, this might be more or less difficult. Either way, new policy creation will need formal approvals from the relevant governing body, and it is sensible to seek input from relevant members of your leadership team to avoid unpicking work at a later stage.

- **Key administrative representatives**, such as those engaged with registration, enrolment, student support, etc.
- **Teachers**. As potentially important end users, they will have useful input on how learning analytics might best inform their practice, and they can flag any immediate shortcomings and/or missing issues.
- **Researchers**. Researchers may feel that their uses of data are covered by other existing policies. It is worth checking, though, and ensuring that all stakeholders are clear on the differences between seeking approvals for data for research purposes and accessing data for operational purposes.

6. Develop the content

It may be tempting to create a report of the process you've gone through, but be mindful that a policy should contain content that is relevant and digestible. Keep it succinct where possible, and set broad aims and objectives. Unless unavoidable, try not to identify specific technologies, as these are likely to evolve as your analytics capability develops. On the other hand, those issues used to define the scope of activities should be retained and made clear. It is perfectly acceptable to refer to other documents that may contain more detail, so long as these are also accessible to pertinent readers.

7. Gain approval

Revisit your key stakeholders and check that your proposals meet with approval. You may find that something emerges at this stage that causes you to revisit an aspect or an assumption; most policy development is an iterative process. At some point, you will need to be prepared to move forward without satisfying every individual. Having said that, your policy must have agreement from your leadership team, and if you are unable to satisfy everyone, be clear why you can't. For instance, you may find that you have conflicting interests between stakeholders; you must be prepared to take a stance and agree on a way forward that meets the agreed purposes and scope, and also satisfies formal or legislative requirements.

Make sure that your written policy is set up in line with your organisational requirements and, if needed, get an understanding of the formal process for approvals.

8. Regular reviews

Policy documents are not static. As might be expected in a fast-moving field, learning analytics will continue to evolve, and activities and approaches that may have been considered unlikely or out of scope may need to be considered in the future. New legislation may provoke the need for review (such as the introduction of the General Data Protection Regulation (GDPR) in the European Union). As with all organisational policies, aim to ensure that your shiny new learning analytics policy is reviewed and updated regularly.

9. Implementation

Development of your policy will no doubt have thrown up a number of issues, and you should consider how these are resolved. If, for example, you have committed to the users of learning analytics being able to understand and interpret learning analytics, you may need to implement some kind of data interpretation training. Creating your policy will be an achievement, but remember, it is of little use if all of those affected by uses of learning analytics in your institution are unaware of it. Promote it in ways that maximise the likelihood of engagement with it. This may mean creating user-friendly versions with links to fuller, more formal versions, creating video content for websites, running workshops, etc. Be creative! The word “policy” is unlikely to generate much excitement, so think of ways that you can entice stakeholders to read and understand your carefully crafted words.

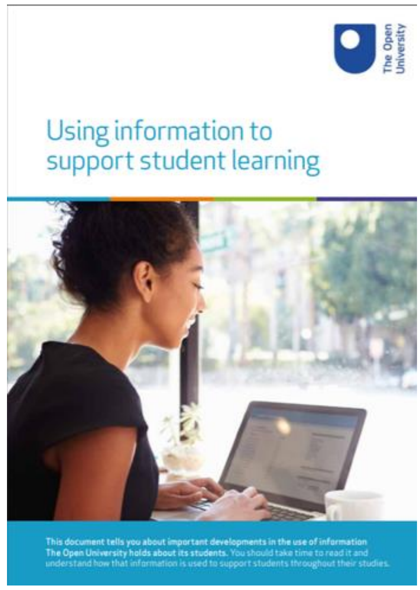
Examples of existing policy and guidance

The first formal policy for educational uses of learning analytics was created in 2014 at The Open University (OU) in the UK, and it is well established as a model on which many subsequent institutional policies have been developed. The process of creating the policy took just under two years, drawing upon published research, legislation and significant input from several stakeholder groups. The OU's policy is built around eight key principles that reflect its open, distance learning ethos:

1. Learning analytics is an ethical practice that should align with core organisational principles, such as open entry to undergraduate-level study.
2. The OU has a responsibility to all stakeholders to use and extract meaning from student data for the benefit of students where feasible.
3. Students should not be wholly defined by their visible data or our interpretation of that data.
4. The purpose and the boundaries regarding the use of learning analytics should be well defined and visible.
5. The university is transparent regarding data collection and will provide students with the opportunity to update their own data and consent agreements at regular intervals.
6. Students should be engaged as active agents in the implementation of learning analytics (e.g., informed consent, personalised learning paths, interventions).
7. Modelling and interventions based on analysis of data should be sound and free from bias.

8. Adoption of learning analytics within the OU requires broad acceptance of the values and benefits (organisational culture) and the development of appropriate skills across the organisation.

An overview of the OU's policy is available at:



https://help.open.ac.uk/documents/policies/ethical-use-of-student-data/files/24/using-information-to-support-student-learning.pdf [43]

Other useful frameworks include the UK JISC's Code of Practice (2015) [44]; the EU-funded DELICATE checklist, produced by the Learning Analytics Community Exchange (LACE, 2016); [45] and the University of Wollogong's policy (2018). [46] As mentioned in Unit 8, the SHEILA framework [47] includes very helpful guidance for reviewing your institution's capacity to adopt learning analytics, and we recommend that you review each of dimensions included to ensure you have considered the most important factors.

[43] A full copy of the policy with greater detail of, for example, included and excluded data sets can be found at: https://help.open.ac.uk/documents/policies/ethical-use-of-student-data/files/22/ethical-use-of-student-data-policy.pdf

[44] https://www.jisc.ac.uk/guides/code-of-practice-for-learning-analytics

[45] http://www.laceproject.eu/ethics-privacy/

[46] https://documents.uow.edu.au/about/policy/alphalisting/uow242448.html

[47] https://sheilaproject.eu/wp-content/uploads/2018/08/SHEILA-framework_Version-2.pdf

Activity: Complete a template for drafting your own policy

Now that you've had an overview of the steps involved in policy creation, you may be feeling a little daunted by the task ahead. Don't be! Copy and paste the template below into a fresh document. Don't spend too long at this stage. Fill in as many parts of the table as you're able; this may help identify which areas you're already well prepared for, and where further thought or effort is required.

Introduction	*Set the scene: describe the context — i.e., your institution is introducing/has begun to use learning analytics. The creation and adoption of a relevant policy will….*
Definitions	*To ensure clear understanding of terms within the policy, e.g., how do you understand learning analytics, what is a course/learner cohort, etc*
Scope	*What activities/data/learner categories will be included within your policy?* *What is excluded? E.g., teacher analytics?*
Policy team	*Try to identify the names (or roles) of the people helping you to develop the policy. Keep it to five or six at most (you can always consult others on an ad hoc basis). Identify a project resource to support the process.*
Policy champion/owner	*Who has suggested the introduction of new policy? Whose responsibility will it be longer term?*
Relevant policy/legislation	*List any existing institutional policies, and check them out for likely impact/crossover.* *Identify the relevant legislation — e.g., data protection in your national/local context (see the activity in Unit 7).*
Stakeholder groups	*Make a list of all of the individuals/groups you should talk to.*

This should be enough to get you started. Don’t attempt to set out your guiding principles by yourself — they will flow from your discussions with stakeholders and your research.

There are a number of guides to producing policy, depending on the complexity of your needs and the time you have available. You may find this UK Government resource, “Policy Lab in a Day,” [48] helpful to get you started if you’re planning to develop policy within a team.

[48] https://assets.publishing.service.gov.uk/media/5791f90de5274a0da30001a1/Policy_Lab_in_a_day.pdf

Summary and conclusion

This unit has tackled an often forgotten aspect of learning analytics: creating and maintaining a policy (or at the very least, a set of principles) to guide how your institution will engage with learning analytics. Although the language of policy can be very dry, creating a policy will really open your eyes to what you want to use analytics for and where you should tread carefully. Good luck!

Check your progress

a. What is the point of creating an institutional policy for learning analytics (choose all that apply):

   i. It sets out in a formal, transparent way how and why the institution will collect, analyse and use learner data.

   ii. It's a hoop that the institution has to jump through.

   iii. It allows the institution to use student data in whatever way it wants.

   iv. It drives the future uses of learning analytics within the organisation.

   v. It creates a shared understanding for all stakeholders and helps to get each group on board.

b. Is it possible to apply existing policy from another institution? (choose one)

   i. Yes, the issues are fundamentally the same.
   ii. Yes, with consideration of institution-specific issues.
   iii. Yes, with consideration of institution-specific issues and relevant legislation.
   iv. No, every institution is different, and it would be more efficient to start from scratch.

c. What is the most important issue when creating a new policy concerning uses of learning analytics?

   i. what you want to achieve
   ii. the data available to you
   iii. the package your institution has already purchased
   iv. existing relevant legislation
   v. stakeholder views

d. What is the first thing you should do when creating a policy for learning analytics?

   i. come up with a name for the policy
   ii. find an existing learning analytics policy to check out
   iii. define the purposes of learning analytics at your institution
   iv. identify the data sets available
   v. talk to the learning analytics champion and find out what they want most

e. What are the benefits of consulting stakeholder groups? (choose all that apply)

   i. getting a broad range of input from different perspectives
   ii. proving that you have asked people who might object to what you are planning
   iii. getting buy-in from those who might be impacted by uses of learning analytics
   iv. highlighting issues that you may not have thought of

f. Who are the most important stakeholders?

   i. the senior management team
   ii. learning analytics developers/researchers
   iii. IT staff
   iv. learners
   v. teachers

g. Do users of learning analytics in your institution have to take any notice?

   i. Yes, the policy sets out the rules.
   ii. Yes, if the policy supports what they plan to do.
   iii. Not if they are not aware of it.
   iv. No, it is guidance only.

h. Which of the following informs what goes into your policy? (Choose all that apply.)

   i. the software you plan to use/are using
   ii. the learning analytics activities you plan to carry out
   iii. the data available to be collected
   iv. relevant institutional policy and external legislation
   v. examples of existing learning analytics policies
   vi. stakeholder views

i. What must be done to make sure your new policy takes effect?

   i. Tell the policy owner it is complete.
   ii. Publish it on your institution's website.
   iii. Get formal approvals (if applicable).
   iv. Email the stakeholders and send them a copy.

j. When should you update your policy? (choose all that apply)

   i. every year
   ii. when senior managers tell me to
   iii. according to the agreed review process
   iv. when relevant legislation changes
   v. Never

k. You've written your policy. What next? (choose all that apply)

   i. Take a well-earned rest.
   ii. Consider any implications.
   iii. Forget it and move on.
   iv. Publicise.
   v. Review and update regularly.

Section 10

Future Trends in Learning Analytics

Introduction

Welcome to the last unit in this course! Here we attempt to not only look at what the future may hold for learning analytics but, equally importantly, think about the research questions and future pedagogical issues that learning analytics may help us solve or improve upon.

Think back to Unit 2, where we discussed evidence, data and predictions. Remember that list in Unit 2, providing some pointers for the trustworthiness of predictions? When we think about future trends in learning analytics, what immediately comes to mind is the expansion of personalised learning, intelligent and autonomous tutoring and learner support systems, and the applications of machine learning and artificial intelligence. Given that this course is aimed at a fairly general audience, we do not go into great detail when discussing machine learning and artificial intelligence, but provide links to a number of OER offering accessible and helpful information.[49] You may also want to revisit Unit 6.

Learning outcomes

After you have worked through this unit, you should be able to:

- discover the social imaginary pertaining to the use of data, notions of the "data gaze," data colonisation, and individual and institutional responses to "data creep"
- engage with different examples and the possibilities of personalised/personal learning environments and intelligent tutoring systems
- critically examine emerging questions and issues in learning analytics, and compile a wish-list based on your specific institutional/course context

[49] A short overview of these two concepts is available at https://www.oercommons.org/authoring/27895-artificial-intelligence-and-machine-learning.

The ever-expanding data gaze: an introduction

In Unit 2, we engaged with the nature of evidence and data and mentioned that somehow, individuals and organisations inherently trust numbers. When someone says his or her viewpoint is “based on data,” the statement often forecloses any questioning or doubt. The history of our “trust in numbers” is almost as old as humanity — whether referring to the number of soldiers in ancient armies or the first census to determine the size of a particular population.[50] Combine our trust in numbers with the data revolution, the increasing volumes of data, and the variety, velocity and granularity of data, and we can understand claims that having access to data is like discovering gold. In her recent publication, Shoshana Zuboff[51] claims that the economic value of data has allowed a new form of capitalism to emerge: surveillance capitalism. We don’t discuss those claims here, but we affirm that there is a global quest for more data, and that data has economic value — resulting in an increase in surveillance.

David Beer[52] and others[53] say that we need to understand what individuals and organisations believe about data and having access to more data in order to understand why there is such a race for ever more data and analytics. Beer uses the term “data imaginary” in reference to the belief that we will be better, quicker, more efficient, more productive, more intelligent, more strategic and more competitive.

[50] A fascinating book on the history and evolution of our trust in numbers is: Porter, M. (1995). *Trust in numbers. The pursuit of objectivity in science and public life*. Princeton University Press. Also see the more recent book by Beer, D. (2016). *Metric power*. Palgrave McMillan.

[51] Zuboff, S. (2019). *The age of surveillance capitalism*. Public Affairs.

[52] Beer, D. (2019). *The data gaze*. Sage.

[53] See, for example, Williamson, B. (2017). *Big data in education. The digital future of learning, policy and practice*. Sage.

As discussed in Unit 2, organisations seem to have embraced this view also, moving toward evidence-based management with claims that their strategies and operational plans are data led or data driven.

The data industry is expanding exponentially. There is a growing range of actors, organisations and corporations involved in collecting, storing, selling, analysing and profiting or profiteering from data and analytics. These actors aim to offer competitive solutions to individuals and organisations through the collection and analysis of data. The data industry is looking to *expand* the collection of data into areas not yet digitised or where the data have not been available, as well as intensifying and deepening the collection and analysis of data in areas where the data industry is already active. Beer[54] refers to this intensifying search for new forms of data or new data collection areas as "data frontiers." This expansion of surveillance has also been referred to as a continuation of historical colonisation,[55] where individuals' data and lives are commodified and sold to the highest bidder.

Beer furthermore suggests that the data industry markets its analytics services as "*speedy, accessible, revealing, panoramic, prophetic and smart*." [56]

Speedy: when organisations cannot afford to wait for analyses and findings, they may be offered solutions that provide "instant," "at the touch of a button" analyses.

Accessible: Gone are the days when you needed a team of experts to do the analysis for you. The data industry offers software solutions that anyone can access without having the necessary competencies or experience in analysing data.

[54] Williamson (2017).

[55] An interesting short video by Nick Couldry on data colonisation is available on YouTube: https://www.youtube.com/watch?v=5tcK-XIMQqE.

[56] Beer (2019, p. 22).

- ***Revealing*:** These analytical services also promise insights from your data that you could never have imagined, increasing not only your sustainability but also your competitiveness and profitability. There are no secrets left — everything that you ever wanted to know can be revealed.
- ***Panoramic*:** No longer do you have to accept that there are some areas of the lives of customers that you don't have data or insights on. These analytical services dangle the promise of a 360 view of your customers.
- ***Prophetic*:** All about removing/minimising uncertainty, this aspect of the service involves making predictions about what will happen in the future.
- ***Smart*:** Smart analytical services can learn autonomously without the need for human intervention to make sense of the data.

In various places in this course, we have expressed our belief that the collection, analysis and use of learners' data can assist both teachers and institutions to fulfil their fiduciary duties to ensure effective, appropriate *and* ethical teaching and learning. We have no doubt that the contractual obligations between institutions and learners include all aspects relating to the collection, analysis and use of learner data. Having said that, we also are cognisant of increasing concerns around the ever-expanding data gaze in education.

The ever-expanding data gaze in education: cause for concern?

As teaching and learning increasingly move online, it seems that the drive to understand how learners learn has resulted in concerted efforts to access more learner data than ever before, in terms of volume, variety, granularity and velocity. The range of data collection is constantly expanding: learners being videotaped during lectures in order to analyse their expressions and participation (or lack of); sentiment analyses on postings in an online discussion forum; multimodal analyses of learners' sensory data and emotional states; analyses of reflective learning journal entries, and so forth.[57]

[57] See the discussion by Prinsloo, P. (2019). Tracking (un)belonging: At the intersections of human-algorithmic student support. Paper presented at the 9th Pan-Commonwealth Forum (PCF9), Edinburgh, Scotland. http://oasis.col.org/handle/11599/3373

While the quest to understand learning behaviours and journeys is understandable, there are growing concerns from scholars, learners and civil organisations about increasing surveillance in and outside of classrooms. Should any data from the learning journey be out of bounds? As we indicated in Unit 7, while the collection, analysis and use of learner data are warranted to fulfil the institution's contractual and fiduciary duties, this does not constitute permission to annex and colonise *every* part of the learning journey, especially without knowledge or consent.

Our earlier notes about the data imaginary and its associated promise provide some insight into the expansion of the collection, analysis and uses of learner data. As educational institutions, especially public institutions, increasingly face budget constraints and demands for accountability, the sector has become a data frontier ready for annexation and resistance.[58]

Reflection action: What are your concerns about surveillance?

This activity is more reflective than previous ones, and there are no right or wrong answers.

a. Think of a context where you are teaching and need access to learner data and analytics to both help you teach more effectively and appropriately, and support your learners to make more informed choices. Which of the following responses resonate with your own views? (You can choose more than one option.)

    i. I would love to have access to more data and analytics to help me teach more effectively and better support my learners.
    ii. The data I have access to from my class observations, formative assessment and engagements with learners is enough for me to make informed decisions about my teaching and the extent to which I support learners.

[58] See https://opendistanceteachingandlearning.wordpress.com/2016/11/14/decolonising-the-collection-analyses-and-use-of-student-data-a-tentative-explorationproposal/

iii. I understand the value of data and analytics, but I also understand concerns around the increasing surveillance of learners.
iv. While I support concerns about increasing surveillance outside of the educational institution, in my view, educational institutions and teachers have a contractual duty to use whatever data can help them teach and support learners better.
v. There should be safe spaces where learner data will not be collected, and where their behaviours won't be tracked and analysed.
vi. I support the collection and analysis of learner engagement and learning data but not the collection of sensory data, analysis of emotions or tracking of mobile phones.

b. Look back at the discussion on the data imaginary and the promise by the data industry to provide institutions and teachers with analyses that are "speedy, accessible, revealing, panoramic, prophetic and smart." If you were the manager of an educational institution or the head of an academic department, how would you consider such an offer? What would be your concerns? And how might the institution benefit from such a service? Or do the concerns and risks outweigh the benefits?

With the help of a robot: the future of teaching?

In 2019, Neil Selwyn, one of the most prolific and critical scholars in the field of educational technology, published a book with the title *Should Robots Replace Teachers?*[59] As expected, there were numerous mixed responses to his question — see, for example, the reviews from David James[60] and David Longman.[61]

While the question in the title of Selwyn's book is provocative, it may also be misleading, since it leads us to think in binary terms — yes/no. Perhaps a more provocative question, addressed towards the end of the book, is: How do we optimally use the potential of new technologies such as artificial intelligence to improve teaching and learning in ethical and appropriate ways?

[59] Selwyn, N. (2019). *Should robots replace teachers?. AI and the future of education*. John Wiley & Sons.
[60] https://www.tes.com/news/book-review-should-robots-replace-teachers
[61] https://tpea.ac.uk/wp-content/uploads/2019/10/Neil-Selwyn-Should-Robots-Replace-Teachers_.pdf

As teaching and learning are digitised and move increasingly online, the volume, variety, velocity and granularity of available data start to require a unique combination of expertise, knowledge and experience, including programming expertise. Making sense of newly available data, emerging variables, interdependencies and fluidity in data flows is simply out of reach for most teachers. Consider for a moment the amount (plus variety, etc.) of data that are generated in an online classroom with 20 learners over a 12-week period. Now compare that to making sense of the data of hundreds or thousands of learners over a 12-week period! How do teachers make sure that learners are progressing and don't fall behind? How would teachers or an instructional support team notice when certain learners fail to log in, or do not submit a formative assignment?[62] Surely help is needed?

At the intersection of human teachers and robots

Three developments require that we seriously consider the (pros and cons of the) potential of machine learning and artificial intelligence in education.

Firstly, making sense of the volume, velocity, variety and granularity of data is becoming humanly impossible.

Secondly, technological advances now

i. allow us to increase our understanding of the data and spot possible links,
ii. point out correlations and
iii. help us understand causation.

Thirdly, there are growing concerns about systems that can learn and act autonomously — for example, in credit scoring[63] and admissions to educational institutions.[64] So how might we proceed?[65]

[62] Prinsloo (2019).
[63] https://www.cgap.org/blog/algorithm-bias-credit-scoring-whats-inside-black-box
[64] https://futurism.com/ai-bias-black-box
[65] In the text that follows, we used and adapted parts of the text by Prinsloo (2019), cited earlier, and published by the Commonwealth of Learning under an Attribution-ShareAlike 4.0 International (CC BY-SA 4.0) licence. Retrieved from http://oasis.col.org/handle/11599/3373.

John Danaher, an academic based in Ireland, classifies four essential components in human decision making:[66] sensing, processing, acting and learning.

- *Sensing* – collecting data from sources
- *Processing* – organising the collected data into useful chunks/patterns as related to categories, goals or foreseen actions
- *Acting* – using the outcome of the processing to implement a course of action
- *Learning* – the system learns from previous collections/analyses and adapts accordingly

Let's consider for a moment how these components will play out in an educational situation. Think of an educator "seeing" a change in behaviour (e.g., the non-submission of an assignment), "processing" the information (e.g., classifying the learner as at risk of failing and in need of follow-up), "acting" (sending the learner a reminder or query about the non-submission) and "learning" deciding whether the reminder/query makes a difference in the learner's behaviour).

What might happen if/when one or more of these elements are shared or taken over by an algorithm? An algorithm may, for example, "sense," or "notice" the fact that certain learners have not submitted their assignments or not logged on for a period. The algorithm then alerts the educator, who *processes* the information (e.g., what steps to take) and how to act (e.g., make a phone call or send an email). When considering courses with large enrolments, algorithms can be programmed to also *process* the information within set parameters and *act* by sending these identified learners a personalised email offering a range of options and care.

[66] Danaher, J. (2015). *How might algorithms rule our lives? Mapping the logical space of algocracy* [Blog post]. http://philosophicaldisquisitions.blogspot.co.za/2015/06/how-might-algorithms-rule-our-lives.html

This information is then recorded by the lecturer (if s/he sensed, processed and acted), or stored by the algorithm. Whoever/whatever sensed, processed and acted then creates the possibility to learn both from what was done and how the identified learners responded (or not) to this intervention. Figure "An overview of human–algorithmic decision making" illustrates a matrix of different possibilities, ranging from

(i) humans doing all the tasks (sensing, processing, acting and learning; to
(ii) humans sharing any or most of these tasks with an algorithm; to
(iii) the algorithm doing some/most/all of these tasks but being overseen by a human; to (4) the algorithm sensing processing, acting and learning autonomously.

At this stage, please see the “Teaching with robots” video.

Watch Video: https://youtu.be/zpS3_Q_pRkg

Video attribution: “Unit 10: Teaching with Robots” by Commonwealth of Learning (COL) is available under CC BY-SA 4.0.

Figure "An overview of human–algorithmic decision making" was adapted from Danaher[23] and provides an overview of the 256 potentially different options and combinations between human and algorithm, from an exclusively human-driven process to a totally autonomous decision-making system with no human involvement.

	(1) Humans perform the task	(2) Task is shared with algorithms	(3) Algorithms perform task: human supervision	(4) Algorithms perform task: no human input
Seeing	Yes or No?	Yes or No?	Yes or No?	Yes or No?
Processing	Yes or No?	Yes or No?	Yes or No?	Yes or No?
Acting	Yes or No?	Yes or No?	Yes or No?	Yes or No?
Learning	Yes or No?	Yes or No?	Yes or No?	Yes or No?

An overview of human–algorithmic decision making

Let us consider the following possibilities:

Sue is a teacher and responsible for an online class of 100 learners. She is assisted by two support staff, who handle technical inquiries and provide learner support. The course runs over 12 weeks and is sequential — that is, learners must complete all activities for the first week successfully before they can access the content and assessment for the second week. Each week has readings, group discussions and group activities as well as automated multiple-choice self-assessment, and a formative assessment.

Following Option 1 in the first column of Figure “An overview of human–algorithmic decision making” suggests that Sue and her two support staff take full responsibility to “see,” to “process” the information, to “act” and to “learn” from the experience. How might it play out in week 1 of the course? Sue and her support staff check each learner’s login details on a daily basis, whether they participated, what they downloaded and how they are doing in the automated self-assessment(s). Each person takes responsibility for tracking the activities of 30 or so learners, every day, for a period of 12 weeks. They will “see” which learners have not logged in, or have failed the automated multiple-choice questions, and/or didn’t finish watching the instructional video. They will “process” the information every day and decide whether a particular learner who has not logged in or has failed the automated self-assessment needs support or whether their behaviour can be understood and explained and does not constitute a risk. If they do decide that they need to follow up, they must then “act” and send an email and perhaps extra resources to the learner, or enquire whether that learner needs assistance. They will also need to keep track of whether the learner’s behaviour changes, and record any changes for future reference. As such, they will “learn” so that they may respond in a more informed way in future.

You no doubt agree that this level of responsibility and care over a 12-week period with 100 learners is simply not sustainable. So, what if Sue and her team share the responsibility (Option 2 in Figure “An overview of human–algorithmic decision making”) and activate an algorithm that takes over the “see” and “process” elements and leaves it to Sue and her team to respond and learn? Or Sue and her team may actually leave the first three elements to the algorithm but keep track of everything and just be responsible for the learning.

In Option 3, the team opts for a pre-programmed algorithm to "look out for" certain behaviours, process the phenomena, respond and learn, while sending daily reports to Sue and her team, who can decide to intervene at any time if they feel that the actions need to be tweaked or revised. The team may even decide to allow the algorithm to take over on a particular week and not get involved (and most probably take a well-deserved break).

So Figure "An overview of human–algorithmic decision making" illustrates a range of possibilities where a pre-programmed algorithmic system can assist or take responsibility for specific elements in the teaching strategy. But as you most probably have thought, there are likely to be some aspects within teaching that humans will be far better at than an algorithm. You may also have thought that there are some aspects where you wouldn't want to run the risk of allowing an algorithm to make autonomous decisions (Option 4). These are crucial points to consider when contemplating how robots, or algorithmic decision-making systems, can assist in teaching.

	(1) Humans perform the task	(2) Task is shared with algorithms	(3) Algorithms perform task: human supervision	(4) Algorithms perform task: no human input
Seeing	Yes or No?	Yes or No?	Yes or No?	Yes or No?
Processing	Yes or No?	Yes or No?	Yes or No?	Yes or No?
Acting	Yes or No?	Yes or No?	Yes or No?	Yes or No?
Learning	Yes or No?	Yes or No?	Yes or No?	Yes or No?

Increasing complexity and ethical considerations

Increasing complexity and ethical considerations

Complexities and ethical considerations in human–algorithmic interactions

The risks and complexities of teacher–robot collaboration are further illustrated in the Figure. You will recognise that Figure “Complexities and ethical considerations in human–algorithmic interactions” has, at its centre, the human–algorithmic decision-making matrix of Figure “An overview of human–algorithmic decision making”. What is new in Figure “Complexities and ethical considerations in human–algorithmic interactions” is the addition of two arrows showing how moving along the “seeing – processing – acting – learning” axis and/or from human control to algorithmic control brings increasing complexity and ethical considerations. You might imagine that “acting” would have more significant implications for a learner’s progression than simply observing or processing the information. Similarly, the complexities and ethical implications inherent when algorithmic decision-making systems act autonomously are well documented (see Prinsloo, 2019) and a major impediment in considering Selwyn’s question of whether robots should replace human teachers.

Examples of algorithmic decision-making systems in education

One of the outcomes of this course is to enable you to engage with different examples and possibilities of personalised/personal learning environments and intelligent tutoring systems.

Whether we think of applied analytics (as discussed in Unit 5) and how the “system” can shape a learning journey to meet particular outcomes, or of intelligent tutoring or chat bots that respond to learners’ queries, all of these examples function toward the right-hand side of Figure "An overview of human–algorithmic decision making", where algorithmic decision-making systems function either with human supervision or autonomously. It is important to remember that *humans* write the initial programs, whether in machine learning or artificial intelligence. Intelligent tutoring systems, chat bots or any form of adaptive analytics all start with humans and data.

Once the algorithm is developed, or the software program is written, then it may have the ability to learn in recursive cycles of data, analysis, and acting on the analysis.

Should you encounter examples of such systems, it will be useful to remember Figures "An overview of human–algorithmic decision making" and "Complexities and ethical considerations in human–algorithmic interactions".

Reflection action: Meet the “robots”

The robots in our midst

The notion of a “robot” conjures up different images — from cute, big eyed, almost child-like figures that walk awkwardly, to fearsome machines of war. Actually, robots are far more ordinary, and often invisible. To illustrate this, we would like you to find a friend or colleague and ask them to join you in this “find the robot” activity!

Choose whatever web browser you use — Google, Firefox, Safari, Chrome and so forth. Once the browser has opened, we would like you to type in the following search term exactly the way we have typed it: robot. Then press “enter.” We would then like you to compare your screens. While there may be the same results, there may also be different results. How do you explain that?

What might be even more interesting is if you get in touch with a family member, friend or colleague in another part of the country or even another country and ask them to do the same. The search results are different because Google, or whatever search engine you use, keeps track of what you have already searched for and has built a profile of what your interests are, your likes and dislikes. When you search for any term, the “robot” in the search engine tries to give you what it thinks will appeal to you most, based on your previous searches.

So the “robot” is already here. Robots are everywhere, and mostly, they are invisible and shaping our lives in ways we may not have imagined.

Exploring “robots” in education

Given that “robots” or algorithmic decision-making systems are already part of our lives, we invite you to look at the following examples of such systems in education. In each example, we list a number of resources and links for you to consider if you have the time or interest.

- Intelligent tutoring systems [67], [68], [69]
- Chatbots [70], [71], [72]
- Adaptive/personalised learning [73], [74], [75]

Case studies

In earlier units, you were introduced to the website of OnTaskLearning[76] and had the opportunity to explore the scenarios (Unit 5) and the tool (Unit 6). We invite you to look at the six case studies shared by OnTaskLearning: UTS, University of South Australia, University of Sydney, UNSW Australia, UTA, and another pilot from the University of South Australia.

The case studies are not long and only briefly introduce the initiative. We invite you to read through the case studies and choose two that, in your opinion, provide an interesting perspective on the potential of using algorithmic decision-making systems in education.

[67] https://medium.com/@roybirobot/how-intelligent-tutoring-systems-are-changing-education-d60327e54dfb
[68] https://www.cmu.edu/news/stories/archives/2020/may/intelligent-tutors.html
[69] https://library.educause.edu/resources/2013/7/7-things-you-should-know-about-intelligent-tutoring-systems
[70] https://www.sciencedirect.com/science/article/pii/S0360131520300622
[71] https://elearningindustry.com/chatbots-in-education-applications-chatbot-technologies
[72] https://medium.com/botsify/how-is-education-industry-being-improved-by-ai-chatbots-4a1be093cdae
[73] https://www.thetechedvocate.org/5-things-know-adaptive-learning/
[74] https://elearningindustry.com/adaptive-learning-for-schools-colleges
[75] https://medium.com/inspired-ideas-prek-12/adaptive-learning-technology-can-change-your-classroom-2b8a2ace8aad
[76] https://www.ontasklearning.org

Towards a research agenda: What questions remain unanswered?

Since the emergence of learning analytics in 2011, the field has grown, expanded and matured. As learning analytics has evolved, many lessons have been learned from tools and interventions that did not work, those that *did* work, other failures, successes and… new questions.

In Unit 3, we explored evidence, data and predictions, and we acknowledged that the latter are very difficult. In this unit, we have touched upon some of the potential of learning analytics that we may see in the near future. More important than predicting the "what's next" in learning analytics, though, is considering key questions that those working in learning analytics may need to ponder.

We conclude this unit, and this course, by inviting you to listen to one of the most prolific and critical scholars in educational technology, Neil Selwyn. In 2017, he was invited to present the keynote at the annual Learning Analytics and Knowledge Conference (LAK17).[77] If you prefer to read his speech rather than watching the video, you are most welcome.[78] The link in the footnote to the written keynote will also allow you to engage with a number of scholars' responses to Neil's keynote.

Looking at or reading the keynote and responses should give you a sense of not only how vibrant and dynamic the field is, but also some of the important issues facing learning analytics.

[77] https://www.youtube.com/watch?v=rsUx19_Vf0Q
[78] https://learning-analytics.info/index.php/JLA/issue/view/464

Reflection action: Towards a wish-list

The final outcome of this unit, and of this course, is phrased as follows:

Critically examine emerging questions or issues in learning analytics, and compile a wish-list based on your specific institutional/course context.

Looking back at the whole course — the content, activities and self-assessments, and illustrations — you are hopefully in a different position than you were before you started!

This last and final activity invites you to reflect on the following:

a. What issues or questions emerged during this course that you want to further explore, read about or find clarity on?

b. In this course, we pointed to the potential, but also some of the risks and/or challenges, in the collection, analysis and uses of learner data. If we ask you to identify at most three aspects of learning analytics that you would want to implement or consider implementing in your own classroom, which three aspects would they be?

c. Thinking about your own institution, what would you like to see put into place at your institution, and what might you do to make it happen?

Conclusion

We sincerely hope this course has opened up learning analytics for you and stimulated your thinking about the collection, analysis and use of learner data. This course is called *Learning Analytics: A Primer*, and we have attempted to provide you with a broad, but hopefully also thorough, overview of learning analytics. We realise that because the course is written for a general audience, some units or sections may have been too basic or too technical, but we have tried to stimulate your thinking!

We conclude here by reminding you and ourselves that learning analytics is about understanding and shaping learning and the contexts in which learning occurs, in effective, appropriate and ethical ways.

Many thanks for joining this course. Do complete the final test to receive a certificate of completion, and please also complete the course feedback questionnaire.

Good luck in your learning analytics journeys!

References

Bharara, S., Sabitha, S., & Bansal, A. (2018). Application of learning analytics using clustering data mining for students' disposition analysis. *Education and Information Technologies, 23*(2), 957–84.

Drachsler, H., & Greller, W. (2012). The pulse of learning analytics understandings and expectations from the stakeholders. In *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge* (pp. 120–9). https://doi.org/10.1145/2330601.2330634

Howard, T. (2014). Journey mapping: A brief overview. *Communication Design Quarterly Review, 2*(3), 10–13.

Khalil, M., & Belokrys, G. (2020). OXALIC: An Open edX Advanced Learning Analytics Tool. In *2020 IEEE Learning with MOOCS Conference* (pp. 185–190). IEEE.

Khalil, M., & Ebner, M. (2015). Learning analytics: principles and constraints. In *EdMedia+ Innovate Learning* (pp. 1789–99). Association for the Advancement of Computing in Education.

Kitchin, R. (2013). Big data and human geography: Opportunities, challenges and risks. *Dialogues in Human Geography, 3*, 262–7. doi:10.1177/2043820613513388

Leitner, P., Khalil, M., & Ebner, M. (2017). Learning analytics in higher education — a literature review. In *Learning analytics: Fundaments, applications, and trends* (pp. 1–23). Springer.

Leitner, P., Khalil, M., & Ebner, M. (2017). Learning analytics in higher education — a literature review. In *Learning analytics: Fundaments, applications, and trends* (pp. 1–23). Springer.

References

Lin, J., Sun, G., Shen, J., Cui, T., Yu, P., Xu, D., ... & Beydoun, G. (2019). Towards the readiness of learning analytics data for micro learning. In *International Conference on Services Computing* (pp. 66–76). Springer.

Prinsloo, P. (2017). *Guidelines on the ethical use of learner data: a draft narrative framework.* https://www.siyaphumelela.org.za/documents/5a61c7b737ff5.pdf

Romero, C., & Ventura, S. (2020). Educational data mining and learning analytics: An updated survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 10*(3), e1355.

Slade, S., Prinsloo, P., & Khalil, M. (2019). Learning analytics at the intersections of student trust, disclosure and benefit. In *Proceedings of the 9th International Conference on Learning Analytics & Knowledge* (pp. 235–44). https://doi.org/10.1145/3303772.3303796

Slater, S., Joksimović, S., Kovanovic, V., Baker, R. S., & Gasevic, D. (2017). Tools for educational data mining: A review. *Journal of Educational and Behavioral Statistics, 42*(1), 85–106.

Authors' bios

Paul Prinsloo is a Research Professor in open and distance learning in the Department of Business Management, the College of Economic and Management Sciences, University of South Africa. Since 2015, he is also a Visiting Professor at the Carl von Ossietzky University of Oldenburg, Germany, and a Fellow of the European Distance and E-Learning Network, and he serves on several editorial boards. Paul is an internationally recognised speaker, scholar and researcher and has published numerous articles in the fields of teaching and learning, student success in distance education contexts, learning analytics, and curriculum development. His current research focuses on the collection, analysis and use of student data in learning analytics, graduate supervision and digital identity. He blogs at https://opendistanceteachingandlearning.wordpress.com, and his Twitter alias is @14prinsp.

Dr Sharon Slade worked until recently as a senior lecturer at The Open University (UK) for 20 years. She was an academic lead for learning analytics projects within the university, creating policy and leading work around ethical uses of student data, the operationalisation of predictive analytics, and approaches for improving retention and progression. She has published around 30 papers on learning analytics, focusing particularly on student consent, perspectives on uses of data, the obligation to act on what is known, the concept of educational triage, and broader issues around an ethics of care. Sharon now works for the Earth Trust, an environmental and educational charity in Oxfordshire in the UK. Her Twitter alias is @sharonslade.

Dr Mohammad Khalil is a senior researcher and lecturer in learning analytics at the Centre for the Science of Learning & Technology in the Faculty of Psychology, University of Bergen, Norway. Mohammad has a master's degree in information security and digital criminology and a PhD from Graz University of Technology in Learning Analytics in MOOCs. He has published over 50 articles on learning analytics in high-standard and well-recognised journals and academic conferences, focusing on understanding and improving student behaviour and engagement in digital learning platforms using data sciences. His current research focuses on learning analytics in open and distance learning, self-regulated learning, mobile technologies, visualisations and gamification, as well as privacy and ethics. Khalil's Twitter alias is @MohdKhalil, and he blogs at http://mohdkhalil.wordpress.com.

Answer to Check Your Progress

UNIT 1

(a) ii , (b) ii , (c) ii , (d) ii , (e) ii , (f) i , (g) ii , (h) i , (i) ii , (j) ii ,
(k) You could have considered the following:

What data do I already have access to that I can use more optimally or appropriately?	What data can my institution provide that will assist me with teaching better and supporting my learners more appropriately?	What data are not available in your institution that, if you had access to them, would help you teach better and support your learners more effectively?	How can you use formative assessment as a strategy to make sense of learners' progress?

UNIT 2

(a) i , (b) i , (c) ii , (d) ii , (e) ii , (f) ii , (g) i , (h) i , (i) iii , (j) iii

UNIT 3

(a) ii , (b) i , (c) i , (d) iii , iv.

UNIT 4

(a) ii, (b) ii , (c) iv , (d) i , (e) iii , (f) i , (g) iii , (h) i , (i) i , (j) iv

UNIT 5

For the answers to questions (a) and (b), you can check your answers against Figure "**An overview of different types of analytics**".

(c).

<table>
<tr><th>Example of questions asked or uses of learning analytics</th><th>Answer</th><th>Which type of analytics best addresses the questions on the left?</th></tr>
<tr><td>I want to provide the learners in my class with a visualisation that will help them to understand their progress and improve their chances of success.</td><td>Diagnostic analytics</td><td rowspan="5">Descriptive analytics
Diagnostic analytics
Predictive analytics
Prescriptive analytics
Applied analytics</td></tr>
<tr><td>I am not sure what is happening in my online class. I see some of the learners are not participating, but I am not sure who is not participating.</td><td>Descriptive analytics</td></tr>
<tr><td>Based on research findings, I have a good idea which of the learners in my class are at risk of failing. I want to send them personalised emails based on their progress and provide them with additional support based on their personal profile.</td><td>Prescriptive analytics</td></tr>
<tr><td>When I look at data of my learners' prior learning experiences, their home language and their previous marks in Mathematics, I want to know which may be at risk of failing Accounting.</td><td>Predictive analytics</td></tr>
<tr><td>Our LMS allows us to adapt each learner's journey according to their profile, previous learning experiences, and learning needs. What type of analytics can I use to actually shape each learner's journey according to their profile?</td><td>Adaptive analytics</td></tr>
</table>

(d) i ,

(e) ii ,

(f) i

UNIT 6

(a) i , (b) ii , (c) iv , (d) ii , (e) ii , (f) ii , (g) iii , (h) iv , (i) i

UNIT 7

(a) 6. None of them. (b) ii , (c) ii , (d) ii, iv

UNIT 8

(a) i , (b) ii , (c) i , (d) iii , (e) i , (f) iv , (g) iv , (h) i , (i) i , (j) iii

UNIT 9

(a) i , iv, v, (b) iii , (c) iv , (d) iii , (e) i ,iii, iv, (f) iv , (g) i , (h) ii , iv, v, vi, (i) iii , (j) ii, iii, iv , (k) ii, iv, v

4710 Kingsway, Suite 2500 Burnaby,
BC V5H 4M2 Canada

Tel: +1.604.775.8200

Fax: +1.604.775.8210

E-mail: info@col.org
Web: www.col.org

June 2021