Learning analytics for language learning and teaching

Hayo Reinders

Unitec, Auckland, New Zealand
info@innovationinteaching.org

If only we could know what our students were up to at any given moment in class. Who is paying attention, and who is falling asleep? Who understands the past perfect and who thinks it is about something wonderful that happened yesterday? And wouldn't it be great if we knew who is motivated and who is ready to drop out of the course? Language teachers perhaps struggle with these questions even more than teachers in other domains, because their students are not able to communicate their preferences and needs as well as L1 speakers. Learning analytics involves monitoring student engagement and comprehension and can be used as a way to identify potential problems early on in a course. In this short article I will describe what learning analytics is, how it can work in practice, as well as its potential benefits and drawbacks for language learning and teaching.

Keywords: Learning Analytics, classroom management, engagement, monitoring, feedback

Learning analytics in action

In a (fictional) community college in Pittsburgh:

It is Monday morning and students fill up the class. As they sit down they use their phones, tablets, and laptops to login to Class Dojo, a popular classroom management program. As each student logs in, Haley, their teacher, can see their picture pop up on the tablet she keeps on her desk. She drags and drops each student's picture to the approximate location where they are sitting in class today. After some informal conversation about what everyone has been up to during the weekend, Haley

Commentary

asks how comfortable they now feel about choosing when to use "should," "have to," and "must" – the topic of the previous class. Each student chooses a number between 1 and 10, with 1 meaning they did not understand it at all, and 10 meaning they understood it perfectly. Haley's screen shows an average of 8.7 for the class; not bad for such a tricky subject. She notices that a couple of students chose 5. She makes a note to follow up with them later.

She can now get on with the business of the day, which is to discuss the project students will be working on this week, in which they need to interview native speakers, members of the community, to ask them what they feel should be done to improve the town's facilities. After a couple of minutes she notices several question marks on her screen and a couple of "turtles" (which students click to indicate she might be going a bit fast). She backtracks a bit and summarises what she's said up till now in different words. The question marks disappear. Every once in a while she will see that a student has posted a message to the rest of the class; just now there was one checking whether one can say "do you must" or whether it should be "must you." Before she can answer that, another student has already answered the question.

Meanwhile, in a (fictional) university language class in Tokyo:

Mike is busy preparing next week's classes when he gets an alert on his phone. Uh-uh, three students in one of the first-year academic writing classes have hit a critical engagement threshold. Previous experiences in this program have shown that students who do not interact a certain amount on the course's online forums are significantly more likely to fail the course. Mike immediately instant messages the students to set up an appointment with them.

Are these students the exception, or is there a potential problem with the course? Mike pulls up all of the chat and forum transcripts to identify how much students interacted outside of the classroom. Mmmhhh...despite a strong start, the frequency of students' posts and the amount of content students post is declining. Mike decides to work with a couple of colleagues to investigate this further. Is it because students have too much work? Is there a clash with another course? Or are the instructions in subsequent lessons not as clear as in the first few weeks? It turns out to be none of these; instead, Mike and his colleagues find a relation between the type and number of posts the teachers make online and student engagement. Too frequent and too directive posts lead to less involvement on the part of the students. Mike decides to pull back a bit on his own posts and takes care to offer more suggestions than instructions.

These two scenarios show how different forms of learning analytics can help teachers to make informed decisions about how to best support learners, in ways that were previously difficult or impossible to achieve. In this article I will briefly describe what learning analytics and related approaches involve, before offering some suggestions on how they can be implemented.

What is learning analytics

Learning analytics (hereafter LA) involves "…an educational application of web analytics aimed at learner profiling, a process of gathering and analyzing details of individual student interactions in online learning activities. The goal is to build better pedagogies, empower active learning, target at-risk student populations, and assess factors affecting completion and student success" (NMC Horizon Project, 2016). As may be clear from this definition, learning analytics can be applied to any domain of learning. In language education, it could

be used to monitor general indicators, such as attendance and performance on tests, as well as specific language-related issues, such as whether students achieve a certain number of target vocabulary items in a certain period of time, or whether particular groups of students struggle with certain grammatical features, more than others. LA can be used to answer everyday pedagogical question that teachers of almost any level of technical skill can apply, as well as to answer broader questions, for example about education systems as a whole, and used by educational policy analysts, ministries of education, and so forth.

What makes LA different from other forms of data about learners that teachers have long used, such as test scores and student feedback, is its sheer volume, its accessibility, and its immediacy. More data is now available about learners, more easily and more quickly than ever before.

LA is closely related to educational data mining (hereafter EDM), which "…deals with the development of methods to explore data originating in an educational context" (Romero & Ventura, 2010, p. 601). Whereas LA is more focused on understanding the whole learning experience holistically, EDM is more inductive and reductionist, looking at individual components of the learning process. Although they are separate disciplines, I will use the two terms together here as I focus mostly on the similarities in their pedagogical implementation rather than on their theoretical differences.

LA and EDM are receiving a lot of attention from researchers, especially as more data is being collected about learners (think of online enrolment systems, used to capture information from students when they first enroll in a program, such as demographic data and information about prior learning) and their learning (such as data collected by learning management systems like Moodle and Blackboard, which are used to provide access to and to manage online resources and communication related to a course; as well as information recorded in apps and on websites). This research is revealing some interesting findings. For example, the Japanese example above shows what researchers at the University of Wollongong in Australia discovered through social network analysis, or an analysis of the types of interaction patterns observed in a particular group. A freely available recent JISC report (2016) includes this and many other such experiences and is well worth reading for the examples of potential applications of LA in education.

However, the benefits of LA are not limited to researchers alone. Many of the practical questions that language teachers face can be investigated with information that is often already available and using tools such as spreadsheets and basic statistical packages, that many teachers are already acquainted with. Below I will describe some of the benefits of learning analytics for language teaching.

Learning analytics for class and program management

As the two scenarios at the start of this article showed, information about student learning can be obtained both during instructional time or outside of it. The first is an example of "synchronous analytics" where information is obtained about students' understanding, level of engagement, task completion, and so on, during a class. The class can be either online or face-to-face; in the latter case it requires students to be connected through a device. For example, a teacher might check students' knowledge of the new vocabulary in the unit at hand and then see how this compares with previous groups. A considerably lower score might indicate that there is some further work to be done, as this group may not have some of the foundational knowledge of other groups.

This type of information is often collected using tools such as classroom management programs, such as the one mentioned in the opening scenario (ClassDojo) or one of many other such tools (including the freely available Google Classroom) that facilitate interactive forms of learning and teaching and that enable the teacher to monitor student engagement and performance. The data they generate and the tools they offer for monitoring learner behaviour allow for immediate, or synchronous, analysis and intervention.

"Asynchronous analytics" is the analysis of information outside of class time. This can be before or after an individual class, or indeed before a course (to predict what will happen with a particular group of enrolled students compared with previous cohorts, based on features of those students such as their previous language learning experiences, their backgrounds, and so on) or after a course (to gain insight into the success of the course, for example). An example would be to identify if there is a relationship between participation in group work in class and development of students' oral proficiency. Showing how the two are related could be a great motivator to engage reluctant students and to increase Willingness to Communicate.

Whereas the former is mostly used for the purposes of classroom management, to make decisions immediately affecting what happens in class, the latter is used for course and program management and usually has a longer time frame.

Pedagogical purpose of learning analytics

Although LA and EDM offer significant potential for language research, I focus here on the benefits of their implementation in a teaching context.

Synchronous analytics

There are a number of potential benefits to using classroom management tools. Administratively speaking, a lot of tedious work can be handled by learning management systems. Who is present and who is not? What percentage of homework exercises did each student complete successfully? In addition, most programs include ways of grading and storing students' work. Of course, the recorded data give insight not just into the learners, but also into the teachers: does the teacher actively engage with the learners? Does she deal with questions and misunderstandings promptly? Clearly, such questions can be helpful to program leaders and other managers.

Pedagogically speaking, such systems allow teachers to monitor student engagement in ways that, especially in large classes, may be difficult to achieve otherwise. For example, during individual or small-group work, the teacher can easily see what every student is working on, who may be lagging, and who is not getting the answers right. This allows for early identification of problems and enables the teacher to intervene and offer help as and when needed.

Similarly, different types of feedback are possible, both from the teacher to the students and from the students to the teacher. Backchannelling, in the context of this type of software, is an example of feedback whereby students can raise questions using electronic means such as through chat or by indicating difficulties (some programs, for example, let users send a question mark or another symbol to the teacher's screen to indicate they have a query or want the teacher to slow down or skip a topic), not just to the teacher but also to other learners, and where this can be done anonymously. The benefits of this type of

communication include lowering learners' anxiety about showing their lack of understanding, recognising that other learners have difficulties too, and the ability to ask questions without having to interrupt the whole class, among others. Such backchannels can also give the teacher a wealth of additional information about the mood of the class, and the overall level of understanding (see Reinders, 2014, for ideas on how to use backchanneling in class).

Another important benefit of learning-related data is that it can be made available to students. For example, students can see how their vocabulary knowledge compares with that of other learners at their level, or they can see how much they spoke in class. They could be asked to reflect on this type of information, to encourage critical thinking and engagement.

Asynchronous analytics

Asynchronous analytics can offer a number of administrative benefits. For example, it allows for a comparison between different groups, different classes, and different programs across semesters or years, as well as between different teachers. This makes reporting for the purposes of compliance and quality assurance much easier, and can also help with applying for funding or additional teaching resources, as quite robust evidence can be gathered. One of the benefits of this type of data is the potential for increased transparency; everything is out in the open, for everyone to see.

There are also significant pedagogical benefits. For example, analysis of prior experiences with certain groups or certain courses may help to identify key moments at which students need to receive more or different support. Analysis of student engagement and performance throughout a course may help with early identification of learning problems and may prompt early intervention.

Another benefit is that through the above mechanisms support can be targeted more precisely, both temporally and individually. Resources can be allocated where they are needed, and feedback can be more customised particular groups or individuals. For example, if students spend considerably more time on exercises dealing with a particular grammatical point, it may be more efficient for the teacher to cover the topic in more detail in class first.

Large sets of data, for example collected over multiple years and multiple cohorts, also allow for the prediction of student performance by identifying particular traits, or factors, either at the student, staff, administrator, or institutional level, that impact on success and retention. For example, findings from work at California State University showed that virtual learning environment (VLE) use explained 25% of the variation in students' final grade and this knowledge was subsequently used to carefully monitor students' online engagement early in the program. Although few studies of this type exist in the language education domain, it is likely that in the near future similar sorts of findings will show how understanding of particular aspects of the target language are related, or are predictors of, other aspects of the language. This type of information can help teachers to sequence their instruction better.

Finally, just as with synchronous analytics, asynchronous data can be made available to students to give insight in to their progress over the duration of a course or an entire program, potentially increasing students' sense of control over their own learning.

Implementing the use of learning analytics

If the idea of using learning analytics has piqued your interest, there are some recommended practices for exploring its potential. Clearly, in most cases, the use of analytics is going to involve people beyond yourself. Although the individual teacher can collect and interpret data from an individual class, it is likely that at the very least some of the questions and concerns you have about your learners, are shared by some of your colleagues. In many cases, it is therefore helpful to open up a discussion with others, to identify common questions/areas for investigation. Similarly, and especially in the beginning, it is unlikely that one person will be able to figure out both the practical and technical aspects of using analytics. Although many free and relatively easy tools are available, it is often useful to work as a team and to think of learning analytics as a form of action research/detective work, where it is helpful to draw on knowledge and skills from different people as well as to cover different areas in the school environment. Some very useful collaborations for example have taken place between language teachers and librarians or self-access/independent learning centre staff. Some questions to start discussion at a staff meeting could include: what are some of the greatest pedagogical challenges faced by your team at the moment? We have observed that … learners do/don't… Have you found the same? How do you measure the effect of any changes that you make to your teaching or language support? What data do you collect about learners' engagement?

On a related note, the use of the learning analytics is likely to involve an investment of time, at least in the beginning. For this reason, it is important to obtain approval from management, who may also be interested to consider any technical, privacy and security issues (see below). Although analytics can be very helpful when employed by individual teachers, deep insights mostly derive from exploring connections between different classes, courses, teachers, resources and methods.

You may notice that I have not yet discussed the technical requirements, or the software needed to carry out analytics. This is because these depend on the type of data you will be collecting. And this depends on the types of questions you want answers to. It is safe to say, however, that in most cases no specialist software is required, possibly beyond a statistical package (although a surprising amount can be done with Excel).

Step 1: What do you want to know?

Before approaching the potentially daunting task of working with different types of data and analyzing these, it is useful to ask yourself what your objectives are for a particular group of learners. The more specific you can make the context, the better. For example you may have a particular issue or challenge with a specific class, or a group of students that you would like to explore. An example that I recently encountered myself was that students on a new course started off very actively engaged in the class, but by about week five many of them started to drop off. As the majority of these learners were online, it was initially difficult for me to find out why, for example by just approaching a few of them informally after class. This gave me a very specific question to investigate. It also helped me to realise that I needed an asynchronous approach, as I was keen to see developments over time and compare these with other groups of students. One thing I was able to do with our school's learning management system was to compare such drop-off rates between my class and those of other teachers' classes. It turned out that absence was usually across the board, and therefore related to issues beyond my class alone. An entirely different situation arose

with a group of learners recently where some were very vocal and taking up most of the available speaking time. I was keen to identify those students who were not proactively participating and so in this case a synchronous approach was needed. I used the few minutes during which students were working independently, to download and copy the chat log into an Excel spreadsheet that I had prepared. This allowed me immediately to visualise the number of turns, as well as a number of words that each individual student had produced. Excel easily allowed me to display this as a buy tried, enabling me quickly to see each individual student's level of engagement.

Step 2: What type of data do you need?

You may not need as much data, or different data from those that are already available to you. Ask yourself what the minimum, easily obtainable amount of data is, and how this can help you to answer your question. In the case of my classroom interaction, I needed to be able to record the number of times each student spoke during class, and the length of time they spoke, something that the classroom management software that I used was able to provide in its logs.

Step 3: How will you analyse the data?

Once you have the data, you will need to decide what kind of analysis you wish to carry out. In some cases this can be done in class, and in some cases it needs to be done outside of class. In the example above, I simply had to download these logs and import them into a spreadsheet for some basic analysis, in order to identify frequency and length of interaction for each learner. With this information in hand, I was able to keep an eye on those students who had been identified as not participating as much. In order to respond more immediately, next I set up a way for the type of data I had analysed to become available to me during the class (again, using Microsoft Excel) by using alerts so that the next time I was automatically warned when a particular student's participation rate dropped below a certain percentage.

In the case of course and/or program management, more detailed and longitudinal analysis may have to be carried out. This may mean importing test results, or class attendance, or simply students answer as to exercises, into a program for analysis, such as a spreadsheet or a statistical package.

A final word here: an easy mistake is to be daunted by the amount of data that is available. Just because the data is there, does not mean you have to investigate all of it. Remember your original question and stick to it.

Step 4: What will you do with the findings?

Consider what you will do with your insights, as well as with the data that you have collected. Are there any implications for other classes, and/or other teachers? How will you disseminate your findings? Could the data be shared with others, or perhaps compared with other classes? This is an area where a collaborative approach is most helpful. For example, by comparing students' performance across cohorts, or by learner differences, such as first language, prior experience in studying English, or self-reported motivation, it may be possible to identify individual learners who will be likely to need additional support, or might benefit from placement in a different group. Again, in my own case, I was able to share my findings with my colleagues, to alert them to the fact that particular groups of students were more likely to drop off in their attendance, in all courses across the program. We have **83**

now decided to set up a small action group, to look more closely at these issues, to decide how we can best tackle them in the future.

Additional considerations:

Clearly, in all of the above you will need to consider the privacy and security of your students and colleagues. Also be careful when implementing this type of approach with your peers or a staff (see the next section). Good and open communication, as early on as possible, is key.

Your Turn

Think of a source of educational data that you have access to. This could be your school's learning management system, students' results from course work, participation in (online) activities, but also data generated from students' activities outside the classroom, such as their use of mobile apps, or online computer games.

Now consider what insights such data might give you. For example, what would it tell you if your students were playing a particular type of online game a lot in their own time?

Consider one pedagogical question you might be able to answer with this data. In the example above, what are the language learning affordances of that type of game? Are there any links between the type of language used in the game environment and the course the students are taking?

Consider one possible downside or danger in using this data. For example, how would the students respond to mixing their personal and educational lives?

Some words of warning

Not everyone is convinced the use of LA is a good idea. For one thing, it is possible that the use of such tools may lead to over-monitoring and micro-managing of students. A deeper concern is that analysis of data can be based on, or encourage, a mechanistic, behaviourist view of learning. The use of synchronous analytics, for example, may encourage the teacher to focus more on metrics and less on the students themselves. There is a growing concern that technology reduces teaching to a number of simple behaviours, each of which can be observed and measured. In practice, most teachers will argue that the complexities of teaching are huge and that much of what they do does not necessarily have an immediately visible impact.

A related issue is teacher accountability. Although all teachers I know want to do the best possible job, many would also have concerns about a "big brother" in the classroom, looking over their every interaction with the students.

Furthermore, there are privacy and security concerns. Where is all this information stored and to whom is it available? How long does it stay "on record"? If a student did poorly in one class because his pet dog died and had a hard time as a result, is he going to be classified as a poor learner for the rest of his school career? Some of these issues are discussed in Polonetsky (2015).

These are serious considerations, which make it all the more important that teachers
84 are aware of the benefits and drawbacks of LA and are actively involved in its use. As

pedagogical experts at the frontline of the learning process, they are the best place to identify which data are worth recording in the first place, and how to best interpret them. As with development of new technologies in society as a whole, it is only through carefully interacting with them, that we can fully understand the drawbacks and benefits. Although there are no straightforward guidelines, perhaps the best advice is to start small, focus on a particular area of interest/problem, and communicate your intentions as early and is widely as possible.

LA, both in its synchronous and its asynchronous forms, offers genuinely exciting opportunities for insights into the language learning process that were previously unattainable. My hope is that teachers will cautiously and proactively embrace these new opportunities.

References

JISC. (2016). *Learning analytics in higher education. A review of UK and international practice.* Retrieved October 19, 2016 from: https://www.jisc.ac.uk/reports/learning-analytics-in-higher-education.

NMC Horizon Project. (2016). *Higher education edition wiki on learning analytics.* Retrieved October 19, 2016, from http://horizon.wiki.nmc.org/Learning+Analytics.

Polonetsky, J., & Tene, O. (2015). Who is reading whom now: Privacy in education from Books to MOOCs. *Vanderbilt Journal of Entertainment & Technology Law, 17* (4), 927–990.

Reinders, H. (2014). Backchannelling in the language classroom. Improving student attention and retention with feedback technologies. *The Journal of Language Teaching and Learning, 4* (2), 84–91.

Romero, C. Ventura, S., Pechenizkiy, M., & Baker, R. (2010). *Handbook of educational data mining.* Boca Raton, *FL: CRC* Press.

Resources

https://classroom.google.com/
Google classroom is a free chrome browser extension and suite of tools for educators.

www.classdojo.com
A classroom management tool, with collaborative features.

www.socrative.com
An alternative to the above

https://www.classcharts.com/
Another alternative

Asynchronous data can be obtained through your learning management system. If your school does not use this, you can download free applications, such as www.moodle.org and Instructure's Canvas https://www.instructure.com/

Analytics tools include software you may already have, such as spreadsheets and statistical tools, such as SPSS (http://www.ibm.com/analytics/us/en/technology/spss/) or the free R

package (https://www.r-project.org/), as well as more specialised tools such as RapidMiner, Weka (www.cs.waikato.ac.nz/ml/weka), KEEL, SNAPP, IBM Cognos, and Ellucian.

If you want to learn more, there is a free MOOC on learning analytics offered by the University of Michigan:

https://www.edx.org/course/practical-learning-analytics-michiganx-plax-0

Some useful journals that publish research in this area include:

Journal of Educational Data Mining, Journal of Learning Analytics, International Conference on Educational Data Mining, Conference on Learning Analytics and Knowledge, International Conference on Artificial Intelligence in Education, ACM *Knowledge Discovery in Databases, International Conference of the Learning Sciences,* and the annual meeting of the American Educational Research Association.

Some good introductory resources include:

Seminal article by Phillip Long and George Siemens:

http://er.educause.edu/articles/2011/9/penetrating-the-fog-analytics-in-learning-and-education

Code of Practice for learning analytics:

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