

# Learning Analytics: The next frontier for computer assisted language learning in big data age

Qinglan Yu

Department of English, North China Electric Power University, Beijing, 102206, China

**Abstract.** Learning analytics (LA) has been applied to various learning environments, though it is quite new in the field of computer assisted language learning (CALL). This article attempts to examine the application of learning analytics in the upcoming big data age. It starts with an introduction and application of learning analytics in other fields, followed by a retrospective review of historical interaction between learning and media in CALL, and a penetrating analysis on why people would go to learning analytics to increase the efficiency of foreign language education. As approved in previous research, new technology, including big data mining and analysis, would inevitably enhance the learning of foreign languages. Potential changes that learning analytics would bring to Chinese foreign language education and researches are also presented in the article.

**Keywords.** learning analytics; big data; CALL; media

## 1 The defining of learning analytics and how does it work

In higher education, businesses and governments, the data focus is increasingly expressed in the term learning analytics. Though still a young concept in education, learning analytics (LA) is frequently used in relation to data clusters. The ubiquity of the term analytics partly contributes to the breadth of meanings attached to it. A reasonable definition of learning analytics comes from the 1<sup>st</sup> International Conference on Learning Analytics and Knowledge. It says, “Learning analytics is the measurement, collection, analysis and reporting of data about learners and their contexts, for purpose of understanding and optimizing learning and the environments in which it occurs.” ( Banff and Alberta, 2011)

Learning analytics, since then, has always been shaping the scope and range of activities in higher education, affecting administration, research, teaching and learning, and support resources. (Siemens and Long, 2011) Broadly speaking, a learning analytics program evaluate large data sets to provide decision makers or teachers with information that can help to improve learning outcomes, such as, grades, retention, or completion. LA collects and analyzes the “digital breadcrumbs” that students leave as they interact with various computer systems to look for correlations between those activities and learning outcomes. A professional LA software could do what the conventional data analysis software could not: it compares a students’ activity with others in the class, with students who previously took the course, even with students who are in other countries but have this class online, to create a model for how each students is likely to fare. In this way, it could capitalize on the vast quantities of data to improve the efficiency of learning.

The most common use of LA is to identify students who appear less likely to succeed academically and to enable or launch targeted interventions to help them achieve better performance. At the same time, some organizations employ LA tools to identify a specific unit, which is generally difficult to students, and make diagnosis analysis. Instructors can then make curricular changes or modify learning activities to improve learning on the part of all students.

The way of LA work is closely related with the massive quantities of big data. Generally, much of the data comes from the learning managements system, including log-in information, rates of participation in specific activities, time students spend browsing online resources, and student's grades. Then, applications that perform data collection and analysis are frequently either built into or added onto the learning management system from which they draw primary data. A lot of learning management system vendors increasingly offer analytics tools tied to their software, while some universities, universities or book publishers design and develop their own LA applications.

LA applications could gather data, analyze the data, generate the reports, and enable interventions. Different applications would dig out different types of data, based on their purpose, but in general, they would pick up data about the frequency with which students access online resources or the grades of the assessments. Similarly, the types of computing and processing might be different, but one approach that most application would perform is the evaluation of previous learner data to create predictive models of successful and at-risk students. The report could show different emphasis, but a very common concern is to show clearly that which types of learners are potentially successful learners. LA application could be very practical because it provide critical information to students, teachers, or other users, to predict their success, or to alert high-risk students to take some actions to avoid the negative outcome.

EDUCAUSE, an American nonprofit membership association created to support those who lead, manage, and use information technology to benefit higher education, has a lot of promotion papers and reports about learning analytics on its official website. According to EDUCAUSE, the earliest for-profit institutions have been among the early adopters of learning analytics, including University of Phoenix, Capella University, and the American Public University System (APUS). The LA software in APUS could pick up data on a daily basis from a massive data warehouse, compares it by using statistical measures and semantic analysis engines, and could work out a student ranking each week, and indicate the potential successful learners, and possible losers, or to indicate if any changes could be made to the current curriculum or learning materials. And researchers in Rio Salado College found that early grades in coursework were not a reliable indicators of student's final grades, therefore, they designed new LA software to record student's log-in frequency, their pace of work, their involvement in discussion forum, and make them more robust indicator.

Other potential areas of LA might involve the construction of learner profile, personalized learning model, social network analysis, concept analysis and affect analysis.

## 2 Learning analytics in big data age

As Thomas L. Friedman argues, "the world is flat". The world seems chaotic everywhere but interlinked anywhere by one common core-the data. People are living in a complex network made up by Internet, computers, mobile devices. Every step and every move would leave some "digital breadcrumbs". It is no long the age when people read textbook or have a hallway conversation and leaves no hint. It is a time, however, when every click of the mouse, every Facebook or Wechat update will leave clear-cut marks. Moreover, with the popularity of online learning, digital students records, sensors, mobile devices now could capture rich data trails and activity streams. People are stepping into a substantial big data age.

Big data is a term used to describe the new context of abundance. The McKinsey Global Institute defines big data as "datasets whose size is beyond the ability of typical database software tools to capture, store, manage and analyze." (James Manyika, 2011) In response to the limitations of existing data-management techniques, a new breed of technologies, databases, and techniques

(e.g., data-mining or knowledge discovery in databases) has been developed. As a consequence, theorists have posited that something fundamental has changed with the data itself, creating a world in which almost all data interactions, including scientific research, are affected. With digital evidence, the age of experience and “gut instinct” is gradually disposed. One notable shift in health care is from clinical practice to evidence-based medicine. Clinical practice focuses on physician’s first-hand experience and their professional instinct, whereas, the evidence-based medicine is based on carefully designed data collection. Medicine is exploring even deeper fields like computational modeling by using simple questions like “who will get sick?” and then acting on the predictions to assist individuals in making lifestyle or health changes. Insurance companies also make the use of predictive modeling to identify the potential high-risk customers. Effective data analysis can produce insight into how personal choices or health habit affect long term life quality. Commercial businesses and companies are beginning their data analytics and make data-driven decisions.

With some overlapping features, big data and learning analytics are closely related to each other. As Diana G. Oblinger, the President of EDUCAUSE indicates: “Analytics marries large data sets, statistical techniques, and predictive modeling. It could be thought of as the practice of mining institutional data to produce ‘actionable intelligence.’” (Diana G. Oblinger, 2007). Virtually, the key emphasis in big data in learning analytics is that the data itself is a point of or a path to value generation in organizations. Data is not simply the byproduct of interactions and activities within an organization. Data is a critical value layer for governments, corporations, and higher education institutions.

Therefore, a lot of universities or institutions launched their professional researches on this area, some universities even designed academic courses to help lead students’ research. For example, in Harvard University, Doctor McLeod Lang have been offering a course, named “Data Science in Education: Big Data, Learning Analytics and the Information Age”. In the class description, it says clearly “Data Science in Education provides an overview of the uses of Big Data in education with the aim of developing students’ ability to critically evaluate the technologies, analyses and consequences of this emerging field.” (McLeod Lang, 2014)

In China, people began to put some emphasis on this research area. Ma Xiaoling and his colleagues try to construct an open and comprehensive LA system, to study the interaction between evolution of time, space and data. (Ma Xiaoling and etc., 2014) More diverse and sophisticated researches on LA comes from the field of instructional technology, where experts like Sang xinmin, Zheng xudong has originated their study from the history and development of learning science, and they have already carefully explored the complicated relationships between learning science and educational technology. (Sang xinmin and etc., 2014) The field of foreign language education has a late start in big data research, although it has done some work in designing smart teaching or learning assistance platforms. In November, 2014, in the 4th Conference of China CALL (computer assisted language learning), Gu Yueguo has announced his ambition in establishing a national Big Data research center with the support of National Social Science Research Funding, specializing big data-related research topics in foreign language education, which eventually mark the beginning of big data and learning analytics research.

### **3 The historical interaction between learning and media in CALL**

This part bases on an overview of the historic debate between learning and media in the development of instructional technology, or say, educational technology. As we see, the development of CALL is a continuous process of borrowing ideas, methods or even terms from the bigger discipline instructional technology.

As for instructional technology, its definition is still in updating. There are two definitions here: the 1994 AECT definition, it goes like this: “Instructional Technology is the theory and practice of design, development, utilization, management and evaluation of processes and resources for

learning”, and AECT 2004 definition. It goes like this, “Educational technology is the study and ethical practice of facilitating learning and improving performance by creating, using, and managing appropriate technological processes and resources.” These two definitions are given by the same organization, the Association of Education, Communication and Technology, but in different years. From both definitions, we know instructional technology is about how to make the best use of technology, including computer and of course, other medias to make the most efficient and effective learning. Therefore, CALL, Computer Assisted Language Learning, is naturally one branch of this major discipline.

As for “media”, people have different ways of defining it. McLuhan is best known for coining the expressions “the medium is the message” (Clark, 1983). It is commonly accepted that media has always played an important role in human civilization. Take Chinese civilization as example, the earliest media of writing starts from non-paper materials, like tortoise shells, cow bone, bronze, bamboo slips, wood, stone tiles, silk cloth, and etc, to paper and printing materials, like books, brochures, newspapers, magazines, and etc. And now we are in the stage of new media, that is, multimedia and hyper-media, like computers, smart-phones, PDAs, tablets and all other portable devices. Public opinions get accumulated and disseminated with unimaginable speed, with the use of social networks, including “Wechat”, “Facebook”, “Twitter”, “You Tube” and etc. The coined word APEC blue originates from Chinese social network, We Chat, means something beautiful but last very short. No one could be sure that whether Longman or Oxford English Dictionary would take APEC blue as a formal entry. Anyhow this word is spread in an amazing speed in certain virtual communities, like We Chat. And this is partly due to the magic of modern information technology, also the magic of the big data, just like every teacher is overwhelmed by terms like, MOOC, Coursera, edX, and Udacity, M-learning and flipped classroom.

The learning and media debate has been on for several decades, and swept around the whole area of instructional technology. This debate start with Clark (Clark, 1983), who claims that instructional designers gains no learning benefits from employing a specific medium to deliver instruction. And any performance or time saving gains researchers observe are the result of uncontrolled instructional method or novelty.

Robert Kozma responded to Clark in his article Learning with Media (Kozma, 1991) by pointing out that Clark’s view of media as “delivery trucks” creates an “unnecessary schism between medium and method”. According to Kozma (Kozma, 1994), media are an integral part of the instructional design process.

Jack Kourmi (1994) also enters the debate and argues that researchers (like Clark) promulgate the “false impression of equipotentiality between media”. He suggests that researchers should move away from “media comparison” and try to develop and refine criteria for pushing media to best effect.

Clark, however, strongly refutes Koumi’s criticism. In a later article (Clark 1994) he repeats his claim that media attributes are not casual in learning-what they are is casual in the cost effectiveness of learning. He emphasizes that media is not an integral part of method and media do indeed have equipotentiality. Clark insists that all methods can be delivered by a variety of media and media attributes, and what designers must choose is the least expensive and the speediest.

Virtually, the great learning and media debate triggered new perspectives and breakthroughs on the development of educational technology. It could carry on because proponents in either camp will likely find fault with the other. However, as Koumi has stated, researchers should worry less about comparisons of media and instead focus their attention on making the best use of each. Taking a deep insight into the progress of CALL, learner needs and new technology has always reacted on each other and their joint work has always been performing a source of evolution and revolution.

Generally, the historical development of CALL could be mainly divided into four periods, the text-based CALL (in 1960s and 1970s), the multimedia-based Call (in 1980s), the net-work based CALL ( in 1990s), and the lately hyper-media based CALL of digital-dominance era (Yueguo Gu, 2006).

It is well known that during the first period, CALL is based on the development of language laboratory, for the focus of audio-lingual approach is on repetition of words, patterns and drills. One of the most prominent projects, costing approximately five million dollars, was the PLATO project, which was designed to provide interactive, self-paced instruction for large number of students. Murphy and Appel had conducted independent assessment on Plato system. The result shows that this system is very appealing to both teachers and students, but work on system does not have consistent positive or negative effect on student achievements or attrition rates. In another word, the use of this large scale, money consuming new media, shows no significant positive influence on learning outcome, which in one way back up Clark's view on media's role as "delivery truck". As for in 1980s, the dawn of communicative language teaching approach, commentators and researchers shifted their emphasis on the debate between learning and media, focusing more on what learners could do with it, that is, the importance of methodology in CALL. Since 1990s, Internet has become available to both learners and teachers, particularly with its wide access to learning materials and resources which are originally confined to certain specialized, expensive hard wares. It is crucial to understand CALL in this period represents more than one homogeneous type of activity. However, network--based CALL is often criticized for a slow access or a broken linkage, or for being messed up because too many learners get access to one website at the same time. Stepping into the 21 century, speed of Internet has been greatly optimized and the new Cloud technology brings both chances and challenges to foreign language learning. With Internet speed as high as 100 Mbytes/second, as well as powerful smart phones, wearable PDAs, Tablets, Portable DVD players, and free WIFI service in most universities campus, the fast-growing hyper media development would inevitably foster the efficiency of learning to an unprecedented stage.

#### **4 The upcoming changes of CALL in big data age**

"The world is flat" and the force of innovation would eventually push away all the blocking force, and occupy every inch of its frontier, as proved in the historical interaction of "learning and debate". Innovations and creations are always born with changes to the existing world. Similarly, the massive amount of data and applications replaced every other tool and give rise to new challenges to all parties of CALL.

Firstly, English teachers and practitioners are no longer knowledge owners or providers, but class organizers and leaders. From "sage on the stage" to "guide on the side", teacher's work load has not been lessened as commonly expected, but even doubled or several times heavier. It is not exaggerating to say that teachers are competing with an invisible force even when they are in the classroom, if you imagine that every student has a smart device at hand, smart phone or ipad, to check the new word either from the text, or the background information of a Shakespearean allusion. The space of interaction has been greatly expanded with various learning platforms, such as Moodle or Blackboard, therefore, teachers are facing great challenges in hiding himself from bullets of all directions.

Secondly, learning could be real learner-oriented, for the learner himself could decide what to learn, when to learn, how to learn, where to learn, how much to learn and how fast to learn. In one word, learning could happen anywhere, anytime. A successful learner should not be a layman of computers or Internet, which is the basis of efficient learning in big data age. Comprehensive knowledge of the course and curriculum, the credits of compulsory or optional course is also essential. One biggest problem that most students find hard to deal with is the self-regulation and self-monitor in autonomous learning. For example, in flipped classroom, a lot of study tasks are separated into smaller chunks and sent to students for their after class learning, and the teacher is expected to help them solves those problems they find in after class autonomous learning. How could every learner finish the demanded work before the deadline? This is especially hard for Chinese students, for most of them got used to teacher-centered classrooms, where everything is previously arranged and they just follow what they are asked to do.

The changes in learning materials could be astonishing, with massive digital materials everywhere, written, audio, visual or virtual. The future textbook would be revolutionized with the essential changes in learning media, and complicated word narrations and descriptions would be replaced by videos or movies, and asking for learner's multiple sensory inputs, including that from eyes, ears, mouths at the same time. The future learners are expected to keep pace with the form and content of MOOCs and the embedded quiz questions.

Finally, the wall of universities would be pushed down to the floor and be open to learners from all walks of life. Classroom learning would be decreased but online learning would be increased; face-to-face lecture would be rare but on line interaction would be more frequent. A learning community would be made up with members from all around the world and the real globalization would not be a dream. Pen and paper test would be gradually replaced by computer or online test, and the computer could record every detail of the test takers performance, including how many times he corrects the answer, and could make it count in his final grade. The assessment could be more correct and objective because computers bear no prejudice to any test taker. And supported by the big data technology, computer will learn how to evaluate writing and place valid score.

As experience and gut feeling are regarded as less reliable compared with decisions made on the basis of massive sets of data, the way of CALL research would experience fundamental changes from the greatest depth. Is random sampling still a good choice when we can get the information of almost all subjects? Is cause-and-effect relationship still worth more concern than correlation between different subjects? Those learner-produced data trails provide valuable insight into what is actually happening in the learning process, and those customer-made data trails suggest ways in which managers could improve their selling strategies. Since people know "what happens" and "how to do", should they always spend valuable research funding and stick to "why does it happen"? What is the value of large-scale research experiment since people could collect massive amount of data on diverse learners, environments, materials, and technologies? What changes will happen soon if the Cloud technology become sophisticated and began to play important roles in the future CALL? The possibilities could be limitless.

## **5 Learning analytics: the new frontier for future CALL**

With MOOC, mini-lecture, and flipped classroom as the keywords of CALL reform, many teachers worried about their future, with the deep apprehension of being disposed away by fast-developing technologies. Therefore, some of them strongly criticize MOOCs (Massive Open Online Courses) for its relatively high cost, limited degree of interaction, and its puzzling identity as the credits from MOOC is hard to be accepted by public universities, especially those in China. However, as Dr. Diana Oblinger, the President of EDUCAUSE points out in its 2014 annual conference, "technology could be the game changer", it is better to embrace all the hyper-media products in this digital-dominance era. And as she emphasized, among the top ten issues of 2014 EDUCAUSE annual report, the No 1 issue is learning analytics, which is fundamentally established on the basis of giant data and complicated data analysis.

The massive amounts of data and applied mathematics replace every other tool that might be brought to bear. With advanced big data tools, the researchers could forget taxonomy, ontology, and psychology. They need not know why people do what they do? The point is they do it, and they can track and measure it with unprecedented fidelity. With enough data, the numbers speak for themselves. With learning analytics, language learners could better understand the learning process, expand their social network, and help them in conceptual development and discourse analysis. And for language practitioners, it is essential to improve teaching, organizational efficiency, and even help to make intelligent curriculum.

When analytics is applied to curricular resources, the traditional view of courses is disrupted. In a real language learning analytics situation, learner's knowledge, attitudes, and skills required can be rendered as a network of relations. The semantic web and linked data are partial instantiations of language concept. Knowledge domain could be mapped, and learner activity can be evaluated in

relation to those maps. Instead of being an “end of course” activity, assessment is performed in real time as learners demonstrate mastery of language skills or concepts. Learning content is not provided in a packaged textbook but is rendered or computed spontaneously, providing each learner with resources relevant to his own profile, his own learning goals, and within his own knowledge domain. In one word, language learning could be flexible and personalized.

Learning analytics is still in the early stages of implementation and experimentation, but it does show great room in its potential development of computer assisted language learning. Numerous questions exist around how analytics relates to existing organizational systems, including privacy, profiling, information sharing and data ownership. Furthermore, since learning analytics risks a return to the traditional behaviorism as a language learning theory if we confine analytics to behavior data, how we can work out meanings more than behavior data?

As proved in government and businesses, learning analytics and big data would play significant role in the future the growing role in the future. And its development in CALL, which always highlights the latest technology, would undoubtedly affirm this trend. Language learning could be a very complex process since it involves psychological, physiological and affect processes. Learning analytics is essential for penetrating the fog that has settled around language learning, and make things simpler to language learners, practitioners and organizations. For teachers, they could predict who could be the “winner” and who could be the “loser”, and to provide earlier interventions. For language learners, the accurate digital analysis could help them to understand their own performance in relation to their peers or about their progress in relation to their personal goals. For policy makers and administrators, who are confronted with tremendous uncertainty in the face of reduced budget and increased global competition, learning analytics could help them to allocate resources, develop competitive advantages, increase the design of intelligent curriculum, and eventually makes successful learning in low cost and high efficiency.

## Acknowledgement

This work is supported by the Fundamental Research Funds for the Central Universities (North China Electric Power University 2014MS80).

## References

1. Banff, Alberta. 1st International Conference on Learning Analytics and Knowledge, 2011.
2. Chris Anderson. The End of Theory: The Data Deluge Makes the Scientific Method Obsolete [EB/OL]. [http://www.wired.com/science/discoveries/magazine/16-07/pb\\_theory](http://www.wired.com/science/discoveries/magazine/16-07/pb_theory). 2008
3. Clark, R. E.. Media Will Never Influence Learning [J]. Educational Technology Research and Development, 1994, 42(2): 21-29.
4. Clark, R.E. Reconsidering Research on Learning from Media [J]. Review of Educational Research, 1983, 53(4): 445-459.
5. Diana G. Oblinger. Academic Analytics: A New Tool for a New Era. EDUCAUSE Review [EB/OL]. 2007. <http://www.educause.edu/library/erm0742>
6. George Siemens and Phil Long. Penetrating the Fog: Analytics in Learning and Education. EDUCAUSE Review [EB/OL] 2011. <http://www.educause.edu/ero/article/penetrating-fog-analytics-learning-and-education>
7. James Manyika. Big Data: The Next Frontier for Innovation, Competition, and Productivity [EB/OL].2011. [http://www.mckinsey.com/mgi/publications/big\\_data/pdfs/MGI\\_big\\_data\\_exec\\_summary.pdf](http://www.mckinsey.com/mgi/publications/big_data/pdfs/MGI_big_data_exec_summary.pdf).
8. Kozma, R. Learning with Media[J]. Review of Educational Research, 1991, 61(2): 179-211.
9. Kozma, R.A Reply: Media and Methods [J].Educational Technology Research and Development, 1994, 42(3): 11-14.
10. Koumi, J. Media Comparison and Development: A Practitioner’s View [J]. British Journal of Educational Technology, 1994, 25(1): 41-57.

11. Ma Xiao ling, Xing Wan li, Feng xiang, Wu Yonghe. Construction of learning analytics system [J]. Journal of East China Normal University(Natural Sc, 2014), 2014(2): 1-19, 39.
12. McLeod Lang, homepage. [EB/OL].2014 <http://www.gse.harvard.edu/course/fall-2014/t510s-data-science-education-big-data-learning-analytics-and-information-age-fall>
13. Sang xinmin and etc,. The vital link between media and learning: a developmental and comparative study of educational technology, Nanjing University Press. 2014.
14. Yueguo, Gu.and Carol Hall. Using Computer in ELT-Technology, Practice and Theory, Foreign Language Education and Research Press. 2006.