

## Guide to identifying Research Questions in Learning Analytics

### 1. Contributors:

*Dr. Geraldine Gray is a lecturer with the Informatics Department at the Institute of Technology Blanchardstown. She specialises in the design and delivery of modules in the areas of data analytics, text analytics, business intelligence, and enterprise application development at levels 7, 8 and 9, and is course coordinator for ITB's online masters in Applied Data Science and Analytics. Geraldine has also supervised a number of post-graduate research students. Prior to joining ITB, Geraldine lectured in IT Tallaght, and also has a number of years of industrial experience developing software for distribution and inventory management. Research interests and peer-reviewed publications include educational data mining and learning analytics, data analytics for computer forensics and analysis of unstructured data.*

*Philip Scanlon is a PhD. researcher at the Insight Centre for data analytics at DCU. His research domain is in the field of Learning Analytics and specifically the analysis of the unique digital footprints created by student interactions with online systems, within a University environment. Philip's background is as an aircraft engineer and data analyst in the aviation industry, which spanned nearly thirty years. During that time he developed an interest in the use of bespoke software solutions in a big data environment and believes data only becomes knowledge when enough questions are asked of it.*

### 2. Introduction

Learning analytics (LA) is a multidisciplinary field with an overall aim of analysing data to better understand, and so improve, the learning experience for each student. "Learning analytics is about collecting traces that learners leave behind and using those traces to improve learning" Erik Duval (<http://www.laceproject.eu/faqs/learning-analytics>).

This data can include demographic information, previous academic history, assignment and exam performance, interactions with library systems, interactions with student support, attendance data, activity on Virtual Learning Environments (VLEs) and use of ICT (Information Communication Technology) resources. However, analysis of historical data alone will not improve learning; data analysis must also inform and evaluate subsequent changes to the learning journey. This can encompass any of the components that combine to influence a learning journey, including campus facilities and services, educators & support staff, and the student.

The following sections look at this data from a number of perspectives, identifying relevant questions and possible areas of research. This is not an exhaustive list of questions, but may serve as a useful resource in brainstorming possible learning analytics initiatives.

### 3. The Data we have

All functions collect, collate and archive data. The most visible data is that collected at institutional level such as Student Information Systems, VLEs like Blackboard or Moodle, library systems (see ORLA's Data Conceptual Model). Individual departments, course boards, or individual staff members may also have local data held on a PC or database



which is less visible as a source of data. Such data can be structured (excel file or database table) or unstructured text (forum posts, assessment submissions, learning resources, emails etc).

Structured data from Student Information Systems (current and prior academic performance; financial and demographic data) facilitates research questions exploring the relevance of this data to understanding factors impacting on learning and learner context.

Structured temporal data from logs of learning related activities (e.g. VLE data) provides data on how students (and academics) interact with campus services and technologies over time. Non-structured data from VLEs and other sources also represents activity over time, and can include forum posts by students, assessments submitted, e-portfolios, and learning resources supplied by academic staff.

#### 4. Sample Research Questions

There is a wealth of data available, and a requirement for more evidenced based research on how best to use this data to improve the learning experience. For example:

- Who is tasked in asking the questions and using the answers to develop a policy in each institution?
- Do institutional staff know what data has been collected?
- Do all staff understand the data and knowledge readily available from VLEs, and how to access it?
  - Is there value in promoting greater awareness and use of this data by staff?
- Local data that is held in a manual or local IT database can be a robust knowledge source, and has often been pre-processed. This data can often be unstructured and not available to the wide institution. How can we document this data, and assess its usefulness?
- Is all relevant information collected within institutions?
- Do we know what information is presently extracted by staff, from VLEs?
- What do staff use the VLEs for and why do some staff not use them?
- Can personally held data be integrated into a networked system?
  - . . . considering issues such as formatting and datatypes.
- Instructors (and so courses) use VLEs in different ways. So the expected behavior will be different for students on different courses. How can an institution wide solution cater for this diversity?
- Is there is single best practice on what information to extract from usage data, or should it be tailored to individual differences and preferences?
- Is there a useful role for predictive modeling, or is descriptive modeling more appropriate?



**ORLA: Online Resource for Learning Analytics** | <http://tinyurl.com/NFORLA>

- What information is most appropriate to give to a course coordinator or tutor, and when; and what information is useful to give directly to the student?
  - How soon in the semester can data provide useful indicators of students that are struggling?
  - To what extent does this vary from student to student, and from course to course?
- Can interaction with the VLE be used to profile students and their potential for completion or dropping out? This source contains vast quantities of structured data and is not used to its potential as its content is not understood by many.
  - Other resources available that record student activity data include library visits, book searches, books checked out and returned. Should this be included?
- What are appropriate time-on-task estimates from activity data?
- Could analysis of student engagement and academic results over time assist in the identification of module change?
- What effective learning dispositions can be inferred from non-structured student generated content, such as language constructs that are indicative of critical thinking used in essays and forum posts; social networks defined by who is replying to whose posts.
- If such analysis is useful, how should it be used to provide constructive feedback to students, in a way that does not encourage students to 'game the system'?
- Is there additional data that should be include in analysis of student engagement? For example data from support departments, social facilities, SU and campus shops.
- Also of relevance:
  - Do we have a legal obligation to provide this information to students?
  - Do we have an ethical obligation to know what is hidden in our data that could make the difference between a student achieving, or not achieving their learning goals?

## 5. The Data we don't have

Learning is a latent variable. HE typically assesses learning via a variety of continuous assessment activities and end of term examinations.

- Do we know what type of learning our assessments and examinations are assessing?
  - Which aspect does the student benefit from (learn and understand), the continuous assessment or the exam process?
- What effective learning dispositions do we want to reward, and what learning dispositions are we actually rewarding?



In addition to data systematically collected, there are other forms of non-structured data that may give rise to a range of interesting research questions. For example, is their scope to extend research on multimodal classroom analytics into the HE classroom? Two interesting examples were presented at the LAK17 keynote by Dr. Sidney D'Mello (available at <https://www.youtube.com/watch?v=3sZmWyhK690>):

- Audio files of classroom activities – this is an anonymous source of data that may provide useful feedback on level of on-task and off-task activities taking place in the classroom. For example:
  - When students are engaged in active learning, is the audio signature different from other classroom noise?
  - To what extent can this data give useful feedback on the classroom experience?
- Video capture to differentiate on-task and off-task behavior from facial analysis.

In addition, the predicted explosion of the Internet of Things and related networks and technologies opens the potential for many areas of novel research such as use of mobile phones to capture ambient noise; gps sensors and attitude sensors to determine context of a location; and use of wearable sensor badges to improve scholastic performance.<sup>1</sup>

## 6. The student perspective

Students come from various backgrounds and are attracted to different types of courses based on many facets of their background and personality. They have chosen their course based on a multitude of reasons including financial, peer influence or expected potential future earnings. Students experiences in a higher education (HE) can vary based on their ability to adapt and integrate into a very new learning and often living environment. Demographics, socio economic background and family academic achievements are all recognized as indicators of a students achievements within the domain of HE.

- To what extent can learning analytics of this data, and subsequent academic engagement data, provide tools to assist in the identification of how students are adapting and coping within their HE environments?
- Is there a benefit from using analytics to inform recommender systems such as course choices, or suitable learning resources?

To date, there has been limited attention paid to the student's perspective on learning analytics. To what extent are we considering the students voice and what they expect from, and would like, from learning analytics. What do students want us to do with their data, and what would they find most useful?

- Do students understand learning analytics? How can their voice and perspective be captured and considered by a learning analytics initiative?
- Should a dedicated module or part of a module provide a forum of support for, and collection of, student specific data?

## 7. Learning analytics for educators



The players in any learning environment include the student themselves, the academic staff teaching each student; the student support staff providing out-of classroom support to each student; and administration staff managing and recording the students journey through the college.

- A lot of research in learning analytics involves the analysis of student data. This leads to the question: is this enough, is student data alone rich enough to give a useful picture of their learning journey? For example, if predictive models of student retention only look at VLE clicks, the assumption is that lecturer activity remains static from year to year, or is unimportant.
- Academic staff get little feedback on their own teaching and the learning resources they produce. Would analysis of data, made available only to the academic involved, be useful in improving the learning experience? If so, what type of data should be analysed, and how should those results be processed.

<sup>1</sup>Watanabe, J. I., Matsuda, S., & Yano, K. (2013, September). Using wearable sensor badges to improve scholastic performance. In *Proceedings of the 2013 ACM conference on Pervasive and ubiquitous computing adjunct publication* (pp. 139-142). ACM.

- Is there useful analysis that can be done on records generated from students' interaction with support and library staff? Are we gathering the right data to do effective analysis?
- Do we need to differentiate between Academic Analytics (analytics to support institutional decision making) v Learning Analytics (analytics to support student learning), and do the relevant people understand the difference?
- Do the relevant people understand what information is being generated and what it can be used for?
- Is it understood what information, and the sources of that information, is used by the decision makers in Faculties/Schools?
- What are other institutions doing, what questions have they asked, did they answer the question?

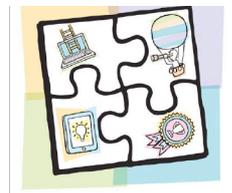
## 8. Ethics

Who:

- Will or should monitor the questions being asked?
- Polices the handling of data usage/storage?
- Sets guidelines for the distribution of information derived from personal data?

## 9. Promoting robust methodologies

There are a number of research questions arising from learning analytics methodologies, encompassing all aspects of a learning analytics project life cycle. There is value in doing some collaborative work in this area, testing various methodologies across a range of student datasets that cover a number of courses and institutions. The purpose would be to test the validity, reliability and robustness of analysis techniques used. Examples



could include:

| <b>Data collection</b>    | <b>Exploring / Preprocessing</b>      | <b>Modelling / Evaluation</b>                                 | <b>General</b>                    |
|---------------------------|---------------------------------------|---|-----------------------------------|
| Impact of sample size     | Handling missing values               | Statistical models: assumptions, p-hacking                    | Justification of method choices   |
| Impact of sampling bias   | Outlier detection                     | Model selection (supervised and unsupervised)                 | Facilitation replication          |
| Defining latent variables | Working with skewed distributions     | Reporting results: model accuracy / model fits / gains scores | Sharing data / data anonymisation |
|                           | Handling class imbalance              | Comparing model accuracies                                    |                                   |
|                           | Valid data transformations            | Unsupervised learning: evaluating results                     |                                   |
|                           | Discretisation / defining class label |   |                                   |

## 10. Conclusion

Learning analytics is a relatively new discipline operating in an ever-increasing digital age that provides ever-richer sources of data. The inherent multidisciplinary nature of research in learning analytics culminates in an exciting variety of research directions<sup>2</sup>. This document offers a brief glimpse of some of these areas including: data we have and data we would like to have; questions that focus on the student, the educators and support staff; ethical and privacy considerations, and promoting robust research methodologies.



<sup>2</sup>C. Lang, G. Siemens, A. Wise, and D. Gavšević, Eds. Handbook of Learning Analytics. 1st ed. SoLAR, 2017. p. 9.  
Available online at: <https://solaresearch.org/hla-17/>