SLATE
Centre for the Science of Learning & Technology

State of the Field Report on Learning Analytics

Kamila Misiejuk & Barbara Wasson

SLATE Report 2017-2

UNIVERSITY OF BERGEN

CENTRE FOR THE SCIENCE OF LEARNING & TECHNOLOGY (SLATE)

VISITING ADDRESS:
Vektergården, Christies gate 13

POST ADDRESS:

Centre for the Science of Learning & Technology (SLATE)
Faculty of Psychology
University of Bergen
PO Box 7807
5020 Bergen

CONTACT:

Kamila Misiejuk (kamila.misiejuk@uib.no)
Barbara Wasson (barbara.wasson@uib.no)

ISBN: 978-82-8088-417-6

TITLE:
State of the Field Report on Learning Analytics

REFERENCE:
Misiejuk, K. & Wasson, B. (2017). State of the Field report on Learning Analytics. SLATE Report 2017-2. Bergen: Centre for the Science of Learning & Technology (SLATE).

FUNDING:
This report is funded through the financial support from the Norwegian Ministry of Education and the University of Bergen.

Table of Contents

Executive Summary

Learning analytics is a young, rapidly growing field of research and practice. In this State of the Field study our goal was to conduct an objective and comprehensive review of learning analytics in order to summarise the field by answering the following questions: What are the main research themes within the field of learning analytics? What data and methods are being used? and What are the characteristics of the learning analytics studies?

The combined proceedings of the Learning Analytics and Knowledge (LAK) conferences from 2011-2015 were used to generate a set of search terms for a systematic search for relevant articles. In collaboration with the Knowledge Centre for Education (Kunnskapssenter for Utdanning) our search produced 796 articles, which after a systematic reduction resulted in a corpus of 100 articles.

A thematic analysis was carried out to identify the primary research themes that have emerged. These included: algorithms and models, data, predictive analysis, learning analytics for educators, learning analytics for institutions, network analysis, tool development, visualisations, overviews, text analysis, and ethics, philosophy & policy.

An analysis of the corpus showed that

- learning analytics is a wide field with articles published in education, computer science, and psychology journals
- the research is data rich, but theory poor
- the majority of the research has been carried out in higher education
- predictive analysis is a very popular research area addressing HE institutional problems such as dropouts, retention, and curriculum issues
- predictive models/algorithms are situation dependent and there is little evidence that they are transferable between different contexts

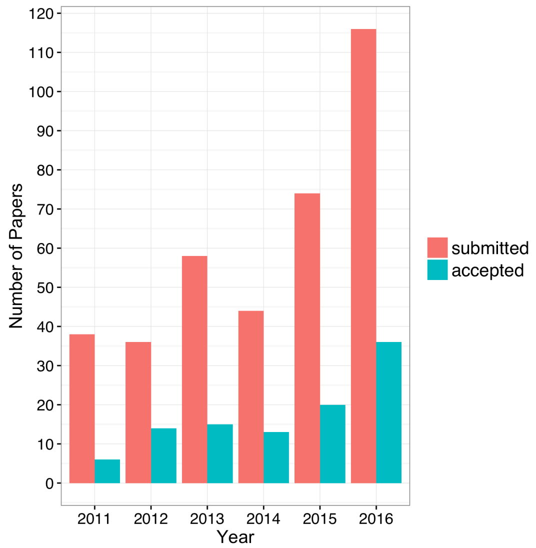
There are also a number of gaps in the research on learning analytics:

- the application of learning analytics in K-12 education (at macro, meso, micro levels)
- research on everyday analytics in classrooms (i.e., how do we collect data in classrooms)
- research on assessment/feedback
- research on learning-centric analytics, as opposed to learner-centric analytics
- Implementation and impact of learning analytics
- data literacy, although there are a few studies addressing whether or not stakeholders can understand the visualisations they are presented.

1. Introduction

Learning Analytics (LA) has emerged over the past 7 years as a promising field of research and domain of practice. Since the term “learning analytics” first started appearing in 2010[1], there has been an increasing number of publications in the area, a growing number of implementations of learning analytics, emerging research centres with learning analytics as a focus, and a growing interest from different stakeholders and policy makers. As the field is in its infancy it is possible to gain an overview of this emerging field by observing the emergence of scholarly societies devoted to the theme.

The Society for Learning Analytics Research[2] (SoLAR) describes itself as an international and interdisciplinary network to support collaborative and open research around learning analytics. Since 2011 the SoLAR community has hosted a yearly conference, Learning Analytics and Knowledge (LAK). Figure 1 shows the submissions and acceptance rates for the LAK conference, with 36 submissions in 2011 to 1316 submissions in 2016.



	ACCEPTED	SUBMITTED	RATE
LAK’11	6	38	16%
LAK’12	14	36	39%
LAK’13	15	58	26%
LAK’14	13	44	30%
LAK’15	20	74	27%
LAK’16	36	116	31%
OVERALL	104	366	29%

Figure 1 Submissions and Acceptance Rates for LAK’11 - LAK’16 (from Misiejuk, 2017)

The most cited definition of the learning analytics comes from the announcement of the 2011 LAK conference:

> LA is the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs.
>
> (Buckingham Shum & Ferguson, 2012, p.4)[3]

The definition tells us that the target of learning analytics includes both learners and learner’s contexts, and the goal of the analysis is not only observation, but also intervention.

[1] http://www.elearnspace.org/blog/2010/08/25/what-are-learning-analytics/

[2] https://solaresearch.org

[3] Buckingham Shum, S. & Ferguson, R. (2012). Social Learning Analytics. *Educational Technology & Society*, 15(3), 3-26.

In July 2013 the first Learning Analytics Summer Institute (LASI) was held at Stanford in Palo Alto. LASI is a summer camp that serves as an intellectual and social springboard to accelerate the maturation of the discipline. Using a tutorial and workshop format, participants can get a flavour of a range of topics and also dive deep into one topic and gain hands-on experience. During the most recent LASI held at the University of Michigan in June 2017, there were 7 workshops and 11 tutorials[4]. Local LASIs have also be arranged around the world, with the first LASI-NORDIC[5] being arranged by SLATE in 2017. This event in Bergen had 45 participants from the Nordic countries and Russia, 2 keynotes, 4 workshops and 17 posters[6].

In 2014 the international peer-reviewed, open access Journal of Learning Analytics[7] was launched at UTS ePress. The journal describes itself as "dedicated to research into the challenges of collecting, analysing and reporting data with the specific intent to improve learning. "Learning" is broadly defined across a range of contexts, including informal learning on the internet, formal academic study in institutions (primary/secondary/tertiary), and workplace learning."[8] and argues that "computational, pedagogical, institutional, policy and social perspectives must be brought into dialogue with each other to ensure that interventions and organisational systems serve the needs of all stakeholders"[9] In the editorial of the inaugural issue[10] they invite research papers and practitioner "hot spots", establishing learning analytics as a field of research and practice. While there were 2 hot spot entries in the first volume, however, there has been no such entries in volumes 2 - 4 (2015-2017).

In order to better understand this rapidly developing field of research and practice, we have carried out a state of the field study of learning analytics and knowledge. In this report we first present a summary of the research in each of these themes, describe the data and methods being used in the research, and characterise the studies. We conclude by summarising the state of the field as a whole, and identifying gaps in research.

[4] http://lasi.solaresearch.org/workshop-list-17/

[5] https://www.slate.uib.no/lasi-nordic2017

[6] Wasson, B. (2017). LASI-Nordic 2017 Posters. SLATE Report 2017-3.

[7] http://learning-analytics.info

[8] http://learning-analytics.info/journals/index.php/JLA/about/history

[9] http://learning-analytics.info/journals/index.php/JLA/about/editorialPolicies#focusAndScope

[10] Gasevic, D., Mirriahi, N., Long, P. & Dawson, S. (2014). Editorial — Inaugural Issue of the Journal of Learning Analytics. *Journal of Learning Analytics,* 1(1), 1-2.

2. A Sense of the Field

We began this research by conducting an exploratory search of the field and its sub domains using the basic search string “learning analytics” in the Web of Science. This gave 369 results, see figure 2, with the first articles appearing in 2010.

Results found:	396
Sum of the Times Cited [?] :	456
Sum of Times Cited without self-citations [?] :	250
Citing Articles [?] :	306
Citing Articles without self-citations [?] :	207
Average Citations per Item [?] :	1.15
h-index [?] :	13

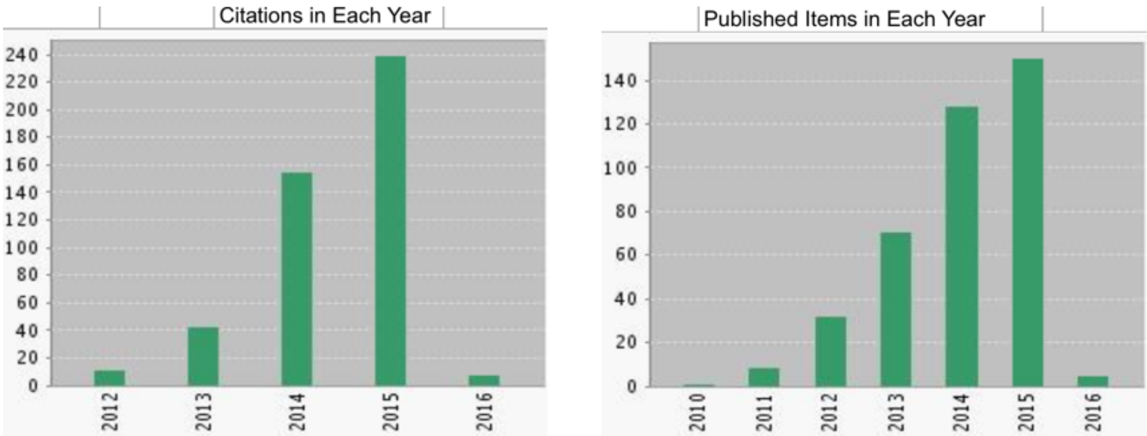


Figure 2 Results from a search on “learning analytics”, Web of Science (February 2016)

The first thing to notice is that the LA articles have relatively low citations (average 1.15), most likely due to the young age of the field. Other characteristics that emerged were 1) there is a wide range of topics covered by the papers, 2) the research is scattered among many disciplines, and 4) many of the most cited papers were, unsurprisingly, overview papers.

In order to identify mainstream LA research, we turned to the Learning Analytics and Knowledge Conference (LAK) proceedings from LAK’11 - LAK’15. In a young field such as learning analytics, the main conference gives a good indication of the breath of the field, those researchers who are central, and the key themes being researched. To get a fast overview of the 264 conference papers, their keywords were visualised in a word cloud, see figure 3. The word cloud shows a wide range of topics, such as computer use in education, to computer-assisted instruction, human factors, measurement, theory, assessment, languages, databases, social network analysis, higher education, decision support, ethics, user interface, etc. The keywords indicated that there are a wide range of topics, a number of disciplines, and a variety of analytics methods and pedagogical approaches being used, confirming the observation of the Web of Science results.

Figure 3 Word cloud of LAK'11 - LAK'15 conference paper keywords.
(The visualisation in higher resolution is in Appendix A)

2.1 Research Questions

Given the diverse research going on within the rapidly growing learning analytics field, we identified the following three research questions to guide our state of the field study:

What are the main themes within the field of learning analytics?

What are the key data and methods are being used?

What are the primary characteristics of the learning analytics studies?

3. Methodology

3.1. Analysis of LAK papers

In order to develop a search string for the formal section of our review analysis the LAK conference keywords and their visualisations and drew on our knowledge of the field by adopting a pedagogical perspective (ignoring and excluding technical issues) and focussing on the implementations of LA in various educational settings. The result was 58 search terms, see figure 4, which can be grouped into five areas:

- problems being addressed (e.g., retention, drop outs, curriculum)
- educational level (e.g, higher education, college, K-12)
- stakeholders (e.g., learning analytics for students, faculties, leaders, rectors, policy makers)
- implementation (e.g., issues such as data management, personalised learning, educational data mining, adaptive learning, learning analytics in MOOCs)
- outcomes (e.g., knowledge building, performance, data literacy, assessment, impact

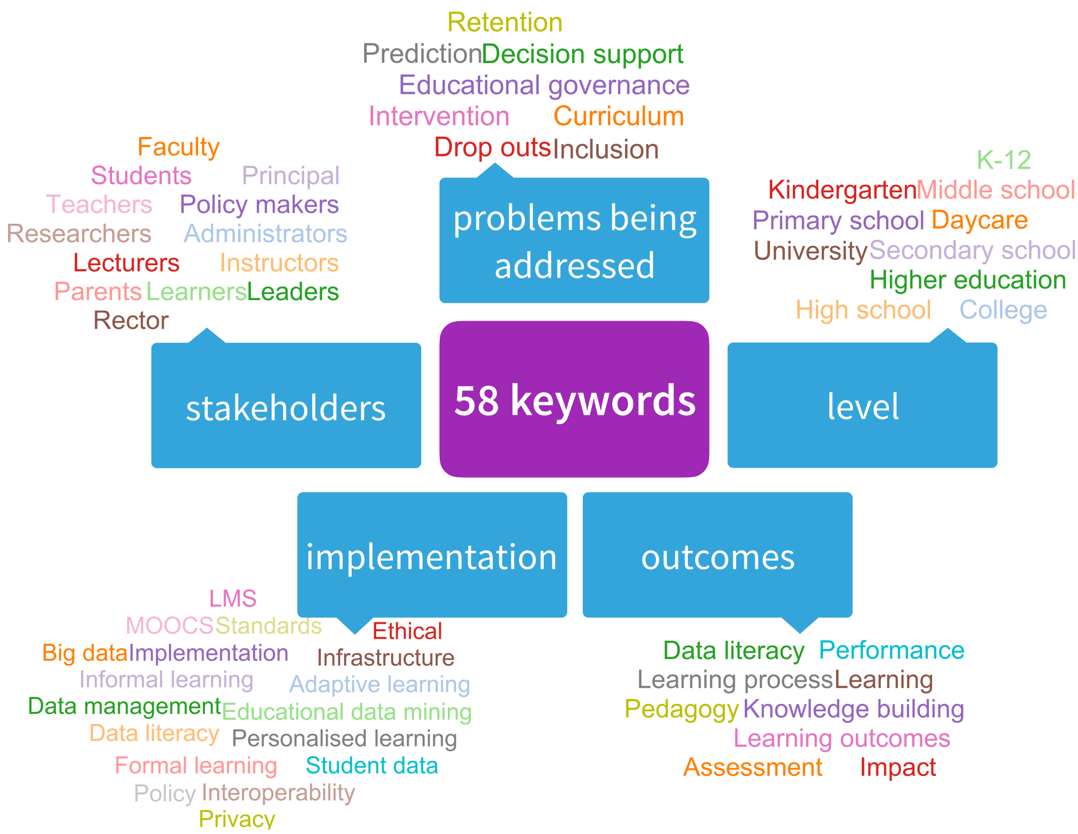


Figure 4 Keywords in the search string

3.2. Search and Sorting

To carry out the search we collaborated with SLATE collaboration partner, the Norwegian Knowledge Centre for Education[11] (Kunnskapssenter for Utdanning), which is specialised in carrying out systematic reviews for the Norwegian Ministry of Education.

Researcher Erik Ruud carried out a search of three electronic databases, ProQuest (including ERIC, PQEJ, ASSIA, IBSS), Scopus, and Psychinfo. Appendix B shows the search string in ProQuest and Scopus format.

The search returned 97, 587, and 112 articles, respectively, from the three databases for a total of 796 articles. The results were imported (title and abstract) to the EPPI-Reviewer 4[12] software, which has been developed for systematic reviews by the EPPI-Centre at University College London. The 161 duplicates were removed, resulting in 635 articles to be sorted. A three-step sorting process, based on pre-determined inclusion criteria was used to prepare the dataset for analysis. Table 1 gives the pre-determined inclusion criteria used for sorting.

Table 1 Inclusion criteria

Inclusion criteria	Explanation
subject	The article should address data and analytics in education or for learning.
study type	The article should be published in a peer-review journal.
research maturity	The research presented in the article should have results of some kind, and not be just speculative.
quality	The article should have a clear research focus/questions, with identifiable research method/design, and the research results are aligned with the research focus/questions.

In first step, the articles were screened for relevance by reading the titles and abstract according to subject and study type. In the second step, the remaining articles were assessed for relevance according to focus/topic, maturity, publishing venue, and quality and a preliminary set of categories was created based on the abstract content of the remaining papers. Finally, in the third step, the full text of the papers was used for a final elimination based on the same inclusion criteria and to create a final categorisation for analysis. Figure 5 shows the flow diagram for the sorting process and the resulting potential categories.

During step 1, Ruud reviewed the titles and abstracts, and eliminated 449 articles that did not meet our inclusion criteria for topic or study type. During step 2, the authors and Ruud reviewed the remaining 185 articles according to our inclusion criteria for focus/topic, maturity, and publication venue, resulting in a further 35 eliminations. The remaining 150 potentially relevant articles were grouped into 5 categories: implementation, impact, learning analytics for …, privacy & ethics, and overview.

In step 3 the authors read the full text of the papers, identified a further 50 papers to be eliminated, resulting in a final data set of 100 articles. While reading we developed new categories, resulting in a

[11] https://www.forskningsradet.no/prognett-kunnskapssenter/KSU/1247146831358

[12] https://eppi.ioe.ac.uk/cms/Default.aspx?tabid=1913

final set of 12 categories: algorithms/models, data, LA+, LA for educators, LA for institutions, network analysis, overview, ethics/philosophy/policy, predictive analysis, text analysis, tool development, and visualisations. It was very challenging to develop the final categories as the dataset is very diverse, making a consistent classification of the papers difficult. We focused on identifying what we understood was the main contribution of the papers, and the result was these categories.

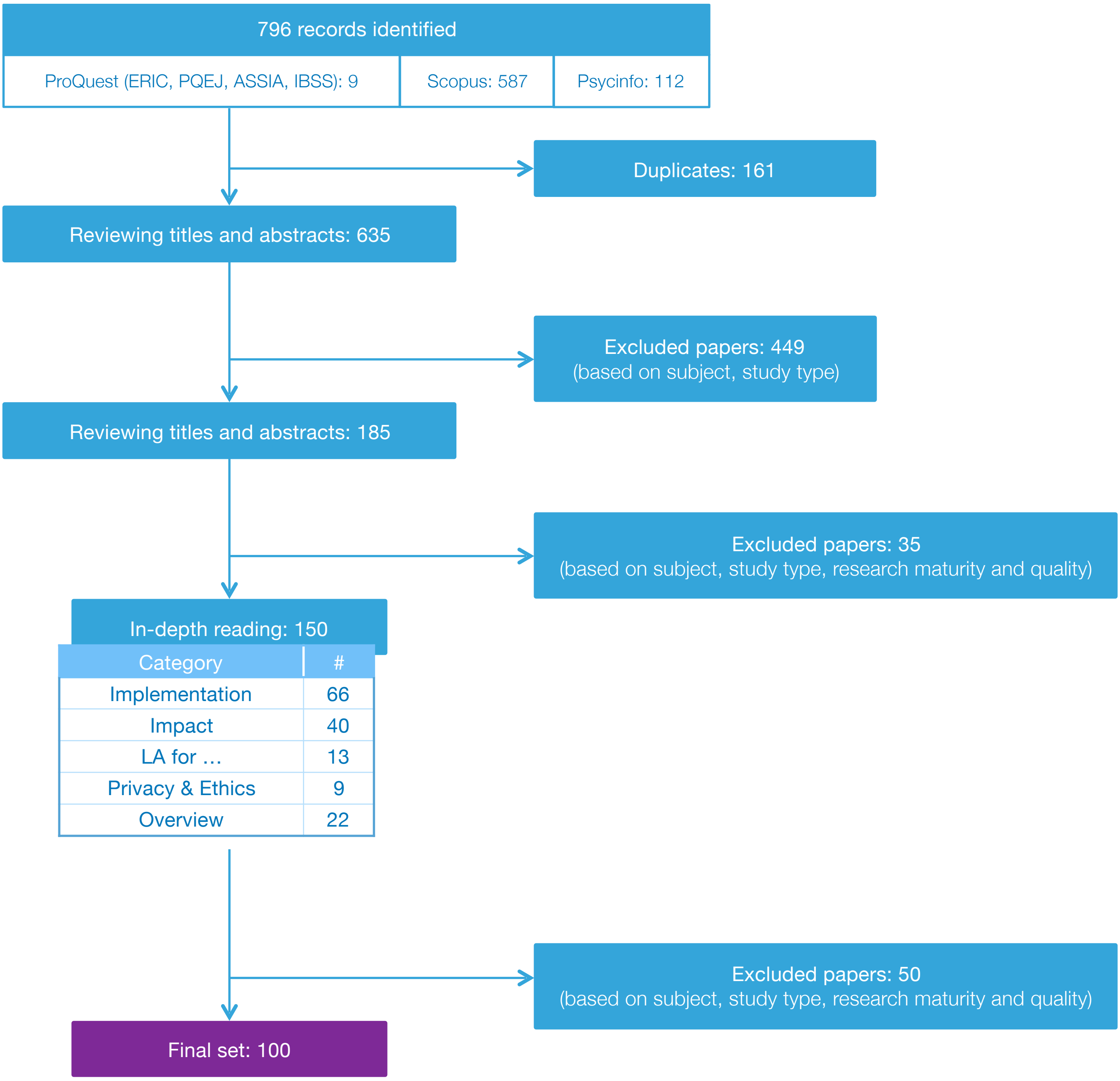


Figure 5 Flow diagram for sorting and the potential theme categories

3. Overview of the Dataset

The final corpus comprises 100 articles. In this section we describe the dataset.

Figure 6 shows the distribution of the papers over publication year, with 15 papers from 2013, 25 from 2014, 40 from 2015, and 20 from 2016 (until February). Figure 7 lists the 43 journals where the articles were published, and shows the distribution among the journals. The most popular journal, *Computers in Human Behavior*, had 13 articles, *British Journal of Educational Technology* had 9, *Journal of Universal Computer Science* had 8, and *American Behavioural Scientist* had 6. The remainder of the journals had between 1 and 5 articles, with 26 journals having only 1 article. The journal titles alone evidence a wide field with education, computer science, and psychology journals being represented.

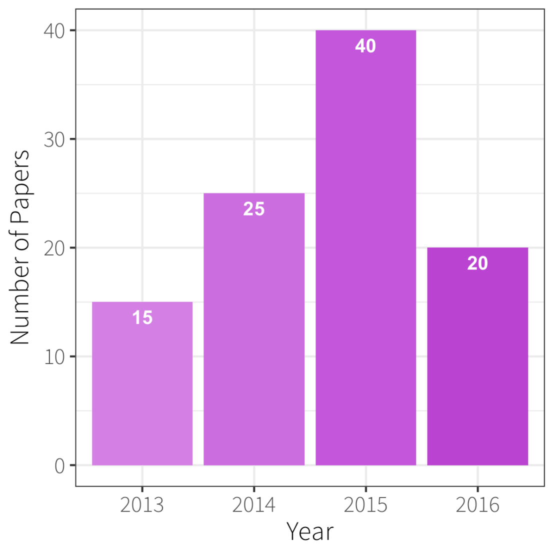


Figure 6 Articles per year

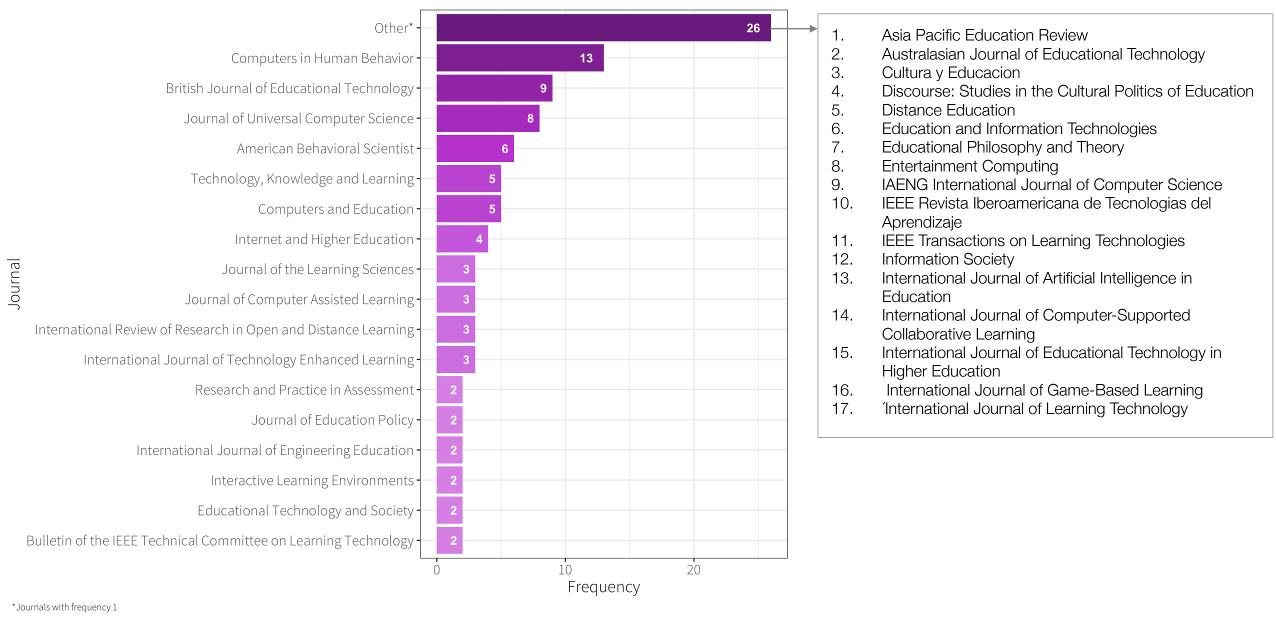
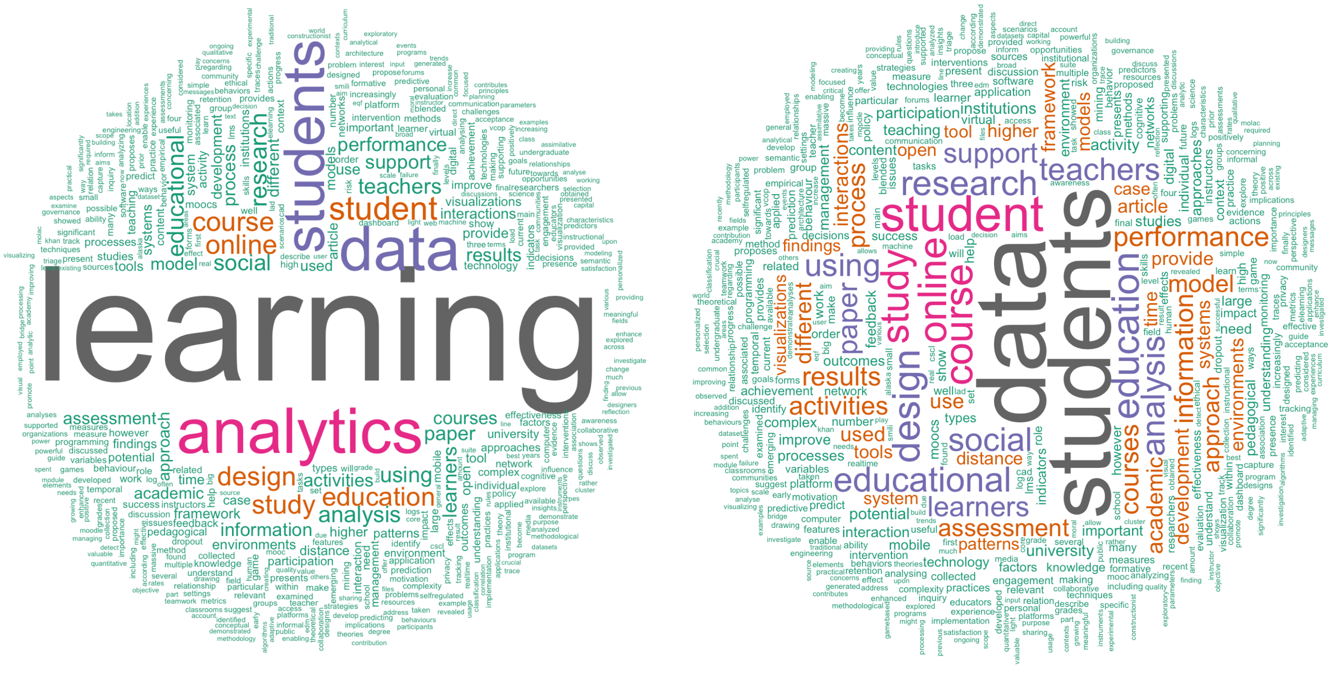


Figure 7 Articles per journal

Figure 8 shows two word clouds of the titles and abstracts of the articles included in the dataset. Figure 8a includes there terms “learning” and “analytics”, and figure 8b does not. See Appendix C for full size figures.



(a) with “learning” and “analytics”

(b) without “learning” and “analytics”

Figure 8 Word clouds of article titles and abstracts

The 20 most common words after learning and analytics, are data, student(s), study, research, design, education, analysis, educational, learners, support, teachers, courses, assessment, activities, results, process, performance, information, model, and academic. As we had a pedagogical and educational focus, these words indicate that our dataset indeed has a focus on educational issues such as performance, assessment, information, and results, teachers and learners, and on process, data, model, analysis, and design.

4. Analysis of the dataset

In addition to being placed in a thematic category, see tables 3-14, each article was analysed according to: contribution kind, data client, educational level, data used, methods used, data subject, pedagogical approach, and learning environment. This section presents our analysis of the dataset and answers the research questions.

4.1 Thematic analysis

In order to answer research question 1

What are the main research themes or directions of research within the field?

we carried out a thematic analysis of the papers that resulted in twelve themes, see table 2. In this section we present a short summary of the articles that fall within each theme.

Table 2 The 12 themes

Category	#	Category	#
Algorithms & Models	12	Network Analysis	8
Data	10	Overview	1
Ethics, Philosophy &Policy	9	Predictive Analysis	20
LA+	5	Text Analysis	3
LA for Educators	10	Tool Development	4
LA for Institutions	11	Visualisations	7

Algorithms & Models

Twelve papers addressed learning analytics algorithms or models.

Three papers, see table 3, presented research on building *algorithms* for understanding aspects of learning processes (individual or group), or predicting student performance. Each of these present work that aims to combine theory and EDM/ML techniques.

Goggins et al. (2015) developed a process-oriented, automatic assessment model for understanding small group learning, which makes complex, small group behaviour visible to teachers via activity analytics visualisations. A web-based tool that uses this algorithm to automatically assess small group learning and visualises the results as time-series activities provides teachers with “actionable intelligence” so they can give real-time support or make interventions with the students. Drawing on methods grounded in complexity theory, the algorithm uses simple interaction rules to model complex small group learning, contributing to a theory-based connection between learning analytics and computation, thus illustrating how theory can inform learning analytics practice.

Table 3 Algorithms & Models articles

	Citation	Contribution Type	Data Client	Educational Setting	Data	Method	Data Subject	Pedagogical Approach	Learning Environment
1	Berland et al. (2013)	Application Model Development	Educators	High School	Log Data	Cluster Analysis	Learners	Constructionist / Constructivist Learning	Immersive Learning Environment Mobile Learning
2	Brooks et al. (2014)	Application Model Development	Educators	University	Log Data Performance Data	Cluster Analysis Correlation Analysis	Learners	-	Blended Learning
3	del Puerto Paule-Ruiz et al. (2015)	Application Model Development	Educators Learners	University	Log Data Performance Data	Association Rules	Learners	Self-directed / Self-regulated Learning	Online Learning
4	Goggins et al. (2015)	Application Model Development Tool Development	Educators	Middle School	Log Data	Cluster Analysis Correlation Analysis	-	Collaborative Learning / CSCL Group Learning	Immersive Learning Environment
5	Joksimovic et al. (2015a)	Application Model Development	Educators Institutions	University	Log Data Performance Data	Correlation Analysis Descriptive Statistics Maximum Likelihood Estimation Regression Analysis T-Test	-	-	OnlineLearning
6	Leony et al. (2015)	Application Model Development	Educators Learners	MOOCs	Log Data	Correlation Analysis Descriptive Statistics	Learners	Adaptive / Personalized Learning	Online Learning
7	Ma et al. (2015)	Application Model Development	Educators	University	Log Data	Structural Equation Modeling	-	Active Learning Collaborative Learning / CSCL	Online Learning
8	Martin et al. (2015)	Application	Learners	Primary School	Log Data Pre- and Posttest	ANOVA Cluster Analysis Descriptive Statistics	Learners	-	Blended Learning Immersive Learning Environment Online Learning
9	Munoz-Merino et al. (2015)	Application Model Development	Educators Institutions	University	Focus Group Data Survey Data	ANOVA Cluster Analaysis Correlation Analysis Data Visualization Descriptive Statistics	Educators Learners	-	Online Learning
10	Nistor et al. (2014)	Application Model Development	Educators	Informal Learning Workplace	Demographic Data Survey Data Text Data	Correlation Analysis Descriptive Statistics Natural Language Processing Network Analysis Regression Analysis	Learners	Social Learning	OnlineLearning
11	Vahdat et al. (2016)	Application Model Development	Educators Learners	University	Survey Data	-	Learners	Adaptive / Personalized Learning Constructionist / Constructivist Learning	-
12	Xing et al. (2015)	Application Model Development	Learners	-	Log Data	Genetic Programming	Learners	Collaborative Learning / CSCL Group Learning	Online Learning

Xing et al. (2015) presented the development of GP-ICRM, a usable prediction model that predicts student performance in a collaborative geometry problem solving environment using a small data set. They used activity theory derived participation indicators as input to a Genetic Programming model, which when combined with EDM and a theory of online participation, results in a prediction model that they postulate is more easily understood by teachers. Using this “practical and interpretable student performance prediction model”, they argue that teachers can discern performance differences in a classroom of students. There is also potential to present the modelling results to students to support learning awareness.

Vahdat et al. (2016) applied machine learning to understand the learning process of humans. Drawing on Cognitive Science and applying Machine Learning to understand human learning, they developed the Human Algorithmic Stability (HAS) algorithm, which measures the capacity of humans to find meaningful rules given various problems in different domains in educational settings. HAS can be used to explain the difficulty level of a particular domain and detect the difficulty level of problems in the domain (i.e., scale the problems from difficult to simple). They suggest that HAS can be integrated as a learning analytics method for personalising and adapting TEL systems to individual students, and to raise awareness of teachers around the difficulty of exercises.

Nine papers addressed *models* for learning analytics for learning process analysis, usage models, prediction of participation, determining procrastination, distinguishing interaction types in online environments, emotional states (boredom, frustration, happiness, confusion), fraction use, effectiveness of resource use, and course prediction.

Berland et al. (2013) explored the opportunities using learning analytics in a constructionist learning environment to *understand the learning progress in a tinkering environment*. The EXTIRE (Explore, Tinker, Refine) model for process analysis of the development of programming skills over time is described. EXTIRE measures the quality of students’ programs over time, and explores the possibility of classifying learners into different clusters based on their learning behaviour.

Brooks et al. (2014) examined the relationship between watching video resources and student performance. Using clustering methods, a usage model is developed and analysed in order to determine if data from *usage of video resources* can predict the overall performance.

Nistor et al. (2014) studied virtual communities of practice (vCoP) and used learning analytics to verify a research model that combines a CoP model and a technology acceptance model that can *predict participation in communities of research*. It is envisaged that the use of such a (combined) model could lead to innovative instructional models and automated tools for supporting vCoPs.

Del Puerto Paule-Ruiz et al. (2015) used association rules to determine which indicators influence *student procrastination*. The model is tested on the data logs from a learning management system (LMS).

Joksimovic et al. (2015a) developed a model to analyse the relationship between the different types of *interaction types* in an online learning environment, academic performance, and the course level. The statistical analysis of the LMS data shows that student-system interactions positively influence academic performance.

Leony et al. (2015) built four models to detect emotions such as *boredom, frustration, happiness, and confusion* in MOOCs. The models were tested on a group of 90 students, and the correlation between the emotions and student’s interaction data was calculated.

Ma et al. (2015) developed a model to evaluate the relationship between student engagement and the role of an instructor in an online learning environment. Eleven hypotheses were tested using structural equation model analysis on data from LMS interactions in 900 university courses. Moreover, the study identified 16 variables (e.g., # instructors in course, # students in course) that influence student engagement in an online environment. The results reveal that course preparation by the instructor influences the students' viewing activities, and guidance and assistance has a significant impact on whether or not students complete learning tasks. In addition, student viewing activities have a positive influence on their completing learning tasks activities

Martin et al. (2015) carried out research on learning fractions. By collecting the usage data from an educational game, a model of *learning fractions by splittin*g was developed. The model was tested on primary school students and the data was analysed using clustering and statistical methods.

Munoz-Merino et al. (2015) attempted to model the *effectiveness with which students use digital learning resources*, and what influences the use. The study provides examples of visualisations of the effectiveness of exercises on students. Moreover, the effectiveness is correlated with other students' behaviour metrics.

Data

Ten papers addressed various aspects of data collection, data analysis, see table 4. A variety of data sources are being used including data logs from games, immersive learning environments, MOOCs, and LMSs, discussion forum messages, interaction data, and observations in face-to-face environments.

Kennedy et al. (2013) developed a prototype that mines, models, and analyses *data from an immersive learning environment* and provides real-time feedback in a 3D immersive surgical simulation. Data about 48 metrics were collected (e.g., current position of drill tool, timestamp, distance of the drill tip) and used to develop Hidden Markov Model topologies of two users groups, novices and experts. Student performance in the simulation is then compared to these topologies in order to determine if feedback should be given (i.e., if they are behaving like a novice). Three important difficulties related to collecting data in learning environments were identified, namely the meaningfulness of the extracted data, the difficulty in providing feedback at the right time, as well as filtering the "noise" in the data.

Thompson et al. (2013) conducted two case studies on the use of learning analytics in collaborative learning scenarios. The first study is of a collaborative learning environment in which users share their nature observations and these *observations are rated* by experts to indicate the user's level of expertise. The second study focused on the *kinds of data* that can be collected and analysed from a *face-to-face collaboration* where a group of four students was given an assignment and their collaboration while solving the task was recorded, analysed, and visualised.

Halverson & Owen (2014) developed a Game-Based Assessment model in an educational game on biology and studied its potential in *capturing data on play and assessment* during game play, in particular focussing on what player interaction data tell about learning in the game. They reported that the most critical issues for learning are progressive successes, and the type of failure that the players experienced.

Table 4 Data articles

	Citation	Contribution Type	Data Client	Educational Setting	Data	Method	Data Subject	Pedagogical Approach	Learning Environment
1	Gibson & de Freitas (2016)	Application Model Development	Learners	University	Demographic Data Focus Group Data Log Data Performance Data	Artificial Neural Networks Association Rules Cluster Analysis Correlation Analysis	Learners	Game-based Learning	Immersive Learning Environment
2	Halverson & Owen (2014)	Application Model Development Tool Development	Educators Learners	Middle School	Log Data Pre- and Posttest	Correlation Analysis Descriptive Statistics	Learners	Drill-based Learning Game-based Learning Interest-driven Learning Play-based Learning	Immersive Learning Environment
3	Kennedy et al. (2013)	Application Tool Development	Learners	University	Log Data	Association Rules Hidden Markov Model	Learners	-	Immersive Learning Environment
4	Liu et al. (2016)	Application	Educators Learners	Middle School	Demographic Data Log Data Performance Data	ANOVA Correlation Analysis Data Visualization Descriptive Statistics	Learners	Collaborative Learning / CSCL Problem-based Learning	Immersive Learning Environment
5	Santos et al. (2015)	Tool Development	Institutions	-	-	-	-	-	-
6	Serrano-Laguna et al. (2014)	Application Tool Development	Educators Learners	-	Log Data	Data Visualization	Learners	Game-based Learning	Immersive Learning Environment
7	Thompson et al. (2013)	Application Model Development	Learners	University	Log Data Observation Data	Descriptive Statistics Discourse Analysis Multimodal Analytics	Learners	Collaborative Learning / CSCL Networked Learning	Face-To-Face Online Learning
8	Xie et al. (2014a)	Application Model Development	Learners	High School	Demographic Data Focus Group Data Log Data Observation Data Survey Data	Correlation Analysis Descriptive Statistics	Learners	Problem-based Learning	Immersive Learning Environment
9	Xie et al. (2014b)	Application Model Development	Learners	High School	Log Data Text Data	Correlation Analysis Descriptive Statistics T-Test Text Mining	Learners	Problem-based Learning	Immersive Learning Environment
10	Yen et al. (2015)	Application Model Development	Educators	University	Log Data Pre- and Posttest Survey Data Text Data	Correlation Analysis Descriptive Statistics Discourse Analysis	Educators Learners	Collaborative Learning / CSCL	Online Learning

Serrano-Laguna et al. (2014) explored the implementation of learning analytics to support assessment for and of learning in educational video games. The first case study found that the application of *data analysis to identify video game design flows* by producing a simple report with heatmap visualisations and graph diagrams, is useful. The second case study focused on *live feedback of student progress* for instructors and on possible interventions to help struggling students. The paper concluded that learning analytics can be a promising tool for improving educational video games.

Yen et al. (2015) examined *Learning Management System (LMS) log data* and attempted to correlate the student's interactions in the LMS with their intrinsic cognitive load, as identified by experts. Moreover, their *discussion forum messages* were analysed and their cognitive load was correlated with their performance, as were synchronous video conferences. Their idea was to provide feedback to students to manage their cognitive load, and for instructors to identify learning problems and assist learners in maintaining a light cognitive load.

Xie et al. (2014a) attempted to identify behaviour patterns in the *data logs from an engineering design program*. Engagement of students over time, gender differences among students, as well as detection of iterative cycles of design were the focus of this study.

Xie et al. (2014b) analysed data from an engineering design project in a digital learning environment in order to assess if data logs can indicate the change of student behaviour after an intervention.

Santos et al. (2015) explored the possibilities of tracking *interaction data* in open learning environments. They presented the design of a learning analytics architecture for collecting and managing learning traces, and describe its implementation.

Gibson & de Freitas (2016) described two learning analytics studies. The first study, the Harvard virtual performance assessment in science, explored the possibility of predicting final assessment grade related to knowledge and skills acquisition by analysing the *data logs from an educational game* taking into consideration *demographic data*. The second study was conducted on a sample of 52,000 university students with 250 records each using a semi-supervised machine learning model. Focus groups helped develop 50 hypotheses about retention and attrition, which would be the first step in developing interventions to help students remain at the university.

Liu et al. (2016) analysed behaviour patterns based on the *log data from an educational game* to understand how the patterns may vary given differences in the learning characteristics of students. Moreover, the relationship between the *student performance* and the log data is analysed, as well as students' engagement levels and fantasy proneness.

Ethics, Philosophy & Policy

There are three articles about ethics, three on philosophy, and three policy articles, see table 5. They addressed the role of the algorithm, analytics as a moral practice, power relations, privacy of digital online information, privacy in relation to autonomy, learning personalisation, and implications of datafication for governance.

Three *ethics* articles indicated the need for system transparency and student control over data. Also from the technical side, there are many concerns about the data management and storage.

Slade & Prinsloo (2013) explored the *power relations* between students and other stakeholders. The starting point of the analysis is the neoliberal consumer-driven market in higher education. Even though the paper comes from a sociocritical perspective, it admits that higher education cannot afford to not use the data.

Table 5 Ethics, Philosophy & Policy articles

	Citation	Contribution Type	Data Client	Educational Setting	Data	Method	Data Subject	Pedagogical Approach	Learning Environment
1	Lundie (2016)	Theory	Researchers	-	-	-	-	-	-
2	Pardo & Siemens (2014)	Theory	Institutions	-	-	-	-	-	-
3	Prinsloo & Slade (2014)	Theory	Institutions	-	-	-	-	-	-
4	Robert-Mahoney et al. (2016)	Theory	Institutions	-	-	-	-	-	-
5	Rubel & Jones (2016)	Theory	Institutions	-	-	-	-	-	-
6	Slade & Prinsloo (2013)	Theory	Institutions	-	-	-	-	-	-
7	Thompson & Cook (2016)	Theory	Institutions	-	-	-	-	-	-
8	Williamson (2015)	Theory	Institutions	-	-	-	-	-	-
9	Williamson (2016)	Theory	Institutions	-	-	-	-	-	-

Table 6 LA+ articles

	Citation	Contribution Type	Data Client	Educational Setting	Data	Method	Data Subject	Pedagogical Approach	Learning Environment
1	Berland et al. (2014)	Theory	Educators	-	-	-	-	Constructionist / Constructivist learning Project-based Learning	-
2	Drachsler & Kalz (2016)	Model Development	Institutions	-	-	-	-	-	-
3	Ellis (2013)	Theory	Researchers	-	-	-	-	-	-
4	Fulantelli et al. (2015)	Model Development Tool Development	Educators Learners	-	-	-	-	-	-
5	Williams (2014)	Theory	Institutions	-	-	-	-	-	-

Pardo & Siemens (2014) framed the discussion of *privacy in higher education* in the context of privacy of digital online information.

Rubel & Jones (2016) described *privacy in relation to autonomy*, as an object of autonomous choice, condition of autonomy, and as limiting other's autonomy.

Three articles can be characterised as *philosophy* articles.

Prinsloo & Slade (2014) presented a more general approach to analytics in education. The focus is on *analytics as a moral practice* and it compared to triage in medicine. Many important questions are discussed such as what is the extent of responsibility to act on knowledge gained from data and what is actually in the best interest of the student.

Thompson & Cook (2016) explored *learning personalisation* in the light of Deleuze's control society. They presented how education, learning, teaching, and belonging are defined differently in the continuous-assessment world.

Lundie (2016) analysed the concept of autonomy within *philosophy of information*. The focus lies on ethical issues around Learning Analytics knowledge in information theory, and the aspect of human learning in contrast to observation of human-computer interaction is emphasised.

Three articles addressed issues related to *policy*.

Williamson (2015) analyses new *challenges and opportunities* in the UK educational system that emerge with widely implementations of digital software and algorithms. A discourse about learning and learners in this new context is presented, especially in context of cross-sectoral intermediary organisations in the English public education.

Williamson (2016) analysed the *potential and challenges of the datafication of the educational system for its governance*. Data visualisation, predictive analytics, and other statistical methods not only open new possibilities for digital governance but also change the nature of education and the basic pedagogical assumptions about learners and learning.

Robert-Mahoney et al. (2016) examined selected US policy papers in order to identify *new trends in thinking about personalised learning* in light of the emergence of learning analytics from an institutional perspective in K-12 education. Some of the findings are that the role of teachers and the definitions of learning and teaching are changing, as well as the growing position of "the algorithm" in contemporary education in the USA.

LA+

The five LA+ articles, see table 6, are more theoretical papers that attempted to "marry" learning analytics with another already established discipline, including assessment, mobile learning, and MOOCs.

Ellis (2013) is a commentary paper, which discussed the potential of *using learning analytics for assessment*. It defines *assessment analytics* and identifies possible application areas.

Berland et al. (2014) investigated the *use of EDM/LA to support quantitative research on constructionist learning*. They saw duality in the relationship between constructionist learning and EDM, the latter having the potential to enhance the ability of constructionist researchers to make rich inferences about learning and learners, while the use of learning analytics raises new research questions and challenges for EDM researchers.

Drachsler & Kalz (2016)[13] described the *potential of learning analytics in the context of MOOCs*. Not only can the MOOC Learning Analytics Innovation Cycle (MOLAC) be applied at the micro, meso, and macro level, but they also highlighted the most important issues that have to be taken into consideration while adapting learning analytics in a MOOCs environment.

Williams (2014) examined the concept of *alternative assessment* (also known as assessment for learning) in the context of learning analytics. Possible implementations of learning analytics on five assessment activities that use either a conventional assessment or alternative assessment approach in learning at scale environments are studied.

Fulantelli et al. (2015) combined l*earning analytics and mobile learning*. Using a task model developed for mobile learning and a Mobile Environment for learning with Linked Open Data (MeLOD), one case study was conducted. It introduced Semantic Web technologies in order to include non-numeric data into the learning analytics analysis.

LA for Educators

Ten articles, see table 7, addressed learning analytics for educators. Issues include whether or not teachers understood the visualisations that they were presented, LA for learning design, curriculum design, detecting low-performing groups, and using LA to support teacher inquiry and provide insights into learning processes.

Florian-Gaviria et al. (2013) examined the use of the adaptive evaluation engine architecture (AEEA) to help teachers use the European qualifications framework in their teaching. AEEA makes use of learning analytics to present an integrated process of modelling, monitoring, and managing lessons to *help teachers understand learner models and learner progress in developing competences*. Results showed that the created visualisations help *teachers understand contextual awareness, kindle reflection, understand students and course progress, and infer patterns of success and failure*.

Lockyer et al. (2013) analyse the possible application of learning analytics in the *learning design process*. They described stages of such implementation using social network analysis to analyse student contributions in a project.

Rodriguez-Triana et al. (2013) examined the last iteration of a project using a design-based research process. The premise is to *help teachers align their pedagogical goals in CSCL situations* with learning analytics and learning design.

Van Leeuwen et al. (2014) described a study in which teachers, divided in experimental and control groups, were shown visualisations and summaries about student's participation and discussions to see if this additional information about student's activities would *influence teacher's interventions and perception about student's performance*. The results showed that teachers and student teachers were *better able to spot the problems regarding participation, intervened more often in problematic groups as time progressed, and displayed more specific explanations of their actions.*

Haya et al. (2015) demonstrated a Social Learning Analytics toolkit that applies social network analysis and content analysis techniques (on forum messages) to analyse collaboration among students and to support teacher inquiry. The research is framed by teacher inquiry and learning design theories. Their results showed that the toolkit supports teachers in *improving the organisation of the learning process* and also supports data that can i*mprove the students' reflection on their own*

[13] 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. 120-129). ACM.

Table 7 LA for Educators articles

	Citation	Contribution Type	Data Client	Educational Setting	Data	Method	Data Subject	Pedagogical Approach	Learning Environment
1	Berland et al. (2015)	Application Tool Development	Educators	High School	Text Data	Descriptive Statistics Semantic Similarity Analysis	Educators Learners	Collaborative Learning / CSCL Constructionist / Constructivist Learning	Immersive Learning Environment
2	Florian-Gaviria et al. (2013)	Model Development Tool Development	Educators	University Workplace	Demographic Data Survey Data	-	Educators	-	Blended Learning Online Learning
3	Haya et al. (2015)	Application Model Development Tool Development	Educators Learners	University	Log Data Text Data	Correlation Analysis Descriptive Statistics Network Analysis Semantic Similarity Analysis	Learners	Active Learning Collaborative Learning / CSCL Networked Learning Social Learning	Online Learning
4	Lockyer et al. (2013)	Theory	Educators	-	-	-	-	-	-
5	McKenney & Mor (2015)	Tool Development	Educators	-	Focus Group Data	-	Learners	-	-
6	Persico & Pozzi (2015)	Theory	Educators	-	-	-	-	-	-
7	Rodriguez-Triana et al. (2013)	Model Development Tool Development	Educators	University	Focus Group Data Log Data	-	Educators Learners	-	Blended Learning
8	Rodriguez-Triana et al. (2015)	Application Model Development Tool Development	Educators	University	Focus Group Data Observation Data	-	Educators Learners	Collaborative Learning / CSCL	Blended Learning
9	Van Leeuwen et al. (2014)	Application	Educators	University	Log Data Survey Data Text Data	Descriptive Statistics T-Test Text Mining	Educators	Collaborative Learning / CSCL	Online Learning
10	Van Leeuwen et al. (2015)	Application Tool Development	Educators	-	Log Data Survey Data Text Data	Descriptive Statistics Regression Analysis T-Test	Learners	Collaborative Learning / CSCL	Online Learning

learning. Furthermore multiple levels of analysis that provides *deeper insights into the collaborative learning process*.

Berland et al. (2015) introduced AMOEBA, a collaboration orchestration tool to support programming learning activities in high school by *supporting teachers to pair students* based on real time analyses of students' programming progressions. The results showed that using AMOEBA to help with pairing students resulted in improvements in the complexity and depth of the student's programs.

McKenney & Mor (2015) attempted to combine learning analytics, learning design, and teacher inquiry in order to *help teachers create teaching resources* that can be shared. The CASCADE-SEA system was used to facilitate the process of curriculum building. The study analysed the results of a user survey on CASCADE-SEA in order to reflect on the possibility of using it as a support tool for teachers.

Persico & Pozzi (2015) argued for, and analysed how learning analytics can support *teacher inquiry and learning design*. It focused on learning design approaches and learning design tools. They argued that learning analytics can "transform learning design from a craft, based on experience, intuition and tacit knowledge, into a mature research area, grounded on data concerning the learning process and hence supporting enquiry while teachers design, run and evaluate the learning process" (p. 230).

Rodriguez-Triana et al. (2015) examined how collaborative learning scenarios can be supported by *learning analytics and learning design*, from the teacher perspective. They connect the pedagogical decisions made at design time with an analysis of the participants' interactions, thus providing teachers with coarse-grained information to help them manage the learning scenarios. The results showed that teachers were positive and that it helped their orchestration of the CSCL scenarios.

Van Leeuwen et al. (2015) explored teacher regulation of CSCL to see if learning analytics tools showing group collaborative activities can *support teachers in detecting low-performing groups*. The results showed that with this support the teachers were not better at detecting problematic groups, but they provided students with more support in general, and in particular, they targeted groups that experienced problems. Two explanations were proposed: 1) the learning analytics steered the teachers' focus towards cognitive activities, and 2) the tools increased the teachers' confidence of their diagnosis.

LA for Institutions

Eleven articles addressed the use of learning analytics for problem solving in institutions, see table 8, such as identifying reasons for dropouts, improving assessment, retention of students, action in a context, and institutional implementation of big data collection and learning analytics. The majority of these papers could fall under academic analytics (Goldstein & Kratz, 2005), but all use the term learning analytics to refer to their research. They also tend to describe studies of the implementation of learning analytics in an institution.

Yasmin (2013) analysed the relationship between students' demographic data, course characteristics, and *dropouts rates* in a University in India. Using a classification tree model, it was determined that students most likely to drop out are either married, have a job, or are over 25 years old. The highest dropout rates were associated with the Mathematics course.

Dawson & Siemens (2014) is a theoretical paper that described new opportunities for learning analytics in the educational system, especially to *improve assessment*. It examined the challenges such as institutional change and introduces a framework for institutional assessment and technology policies.

Table 8 LA for Institutions articles

	Citation	Contribution Type	Data Client	Educational Setting	Data	Method	Data Subject	Pedagogical Approach	Learning Environment
1	Dawson & Siemens (2014)	Model Development Theory	Institutions	-	-	-	-	-	-
2	de Freitas et al. (2015)	Application Model Development	Institutions	University	Demographic Data Log Data Performance Data	Cluster Analysis	Learners	Adaptive / Personalized Learning	Blended Learning
3	de Laat & Prinsen (2014)	Model Development Theory	Institutions	-	-	-	-	-	-
4	Dix & Leavesley (2015)	Model Development	Institutions	-	-	-	-	-	-
5	Ferreira & Andrade (2014)	Model Development Theory	Institutions	-	-	-	-	-	-
6	Macfayden et al. (2014)	Model Development	Institutions	-	-	-	-	-	-
7	Prinsloo et al. (2015)	Theory	Institutions	-	-	-	-	-	-
8	Rienties & Toetenel (2016)	Application Model Development	Institutions	University	Learning Design Data Log Data Performance Data Survey Data	Correlation Analysis Descriptive Statistics Regression Analysis	Learners	Social Learning Constructionist / Constructivist Learning	Online Learning
9	Toetenel & Rienties (2016)	Application	Institutions	University	Learning Design Data	Correlation Analysis Descriptive Statistics	-	Self-directed / Self-regulated Learning	-
10	Volk et al. (2015)	Application Model Development	Institutions	Middle School	Demographic Data Log Data Performance Data	Correlation Analysis Descriptive Statistics Regression Analysis	Learners	Collaborative Learning / CSCL	Online Learning
11	Yasmin (2013)	Application Model Development	Institutions	University	Demographic Data Performance Data	Decision Tree Learning Descriptive Statistics	Learners	Self-directed / Self-regulated Learning	-

de Laat & Prinsen (2014) introduced the concept of Social Learning Analytics (SLA), its potential implementation in higher education, and the associated challenges. SLA uses connectivity and activity and they explore how this can be used in formative assessment practices (rewarding engagement) and in supporting and strengthening *students as active learners* in open and social learning environments. In particular, they argued for moving students focus from awareness to productive engagement in learning activities that promote co–construction of knowledge.

Ferreira & Andrade (2014) showcased an *implementation of learning analytics at the micro, meso, and macro organisational levels at a university*. The implementation process was described on within dimensions: organisational, educational, and technological, and they present a potential Academics Analytics Architecture .

de Freitas et al. (2015) focused on *retention* at a university from the institutional perspective. By analysing the demographic and socioeconomic data, academic results data, student survey data, and data from the LMS, five retention hypotheses were tested. The study found that students are more likely to be retained if they are closer to the average age of their cohorts, engage with the online lecture materials, and are happy with their academic performance. To examine data from a variety of sources, a model was developed that uses both qualitative and quantitative approaches.

Prinsloo et al. (2015) described the potential of learning analytics in the higher education context. Not only the systemic *changes necessary for implementation* and areas of implementation are described, but also a critical look at big data in education is taken. A shift in understanding of data (is more better?), skillsets (from measurement to analytics models), approaches (pattern search), complexities of data (noise, messy), and what is the driving force (of learning and student success) that needs to be considered.

Volk et al. (2015) is a descriptive study that highlights user's behaviour in online language learning environments in higher education. It examined aspects such as time, location, and student performance to better *understand learning processes*.

Dix & Leavesley (2015) are concerned with moving from academic to learning analytics, and argue that LA has to fit in with a busy and fragmented academic life. To this end they developed a model of the learning resource lifecycle based on actors, agents, and events to support the *integration of learning analytics into higher education*. It focused on action-oriented analytics that facilitate thea purpose of *action in a context* and suggests strategies of implementation.

Rientes & Toetenel (2016) analysed the behaviour of students in 151 modules at the Open University, UK, and examine the use of multiple regression models to analyse the *impact of learning design* on student behaviour, satisfaction, and performance. The study finds that learning design is important for producing and understanding VLE behaviour and performance of students. Another important finding was that the number of communication activities is the main predictor of retention.

Toetenel & Rientes (2016) examined the *learning design patterns* in courses taken by 60 000 students at Open University, UK. The courses were mapped according to a learning design taxonomy that reveals that two types of learning activities, assimilative (reading, watching videos, listening to audio) and assessment were the most widely used by the majority of educators. No positive correlation was found between the 7 activity types and student outcomes, however, initial findings suggested that student outcomes are negatively correlated with a high proportion of assimilative activities. They ask that more institutions make their learning decisions explicit and make data available to validate their findings.

Macfayden et al. (2014) discussed that how in order to take advantage of the potential benefits of learning analytics — self-regulated learning, student success — higher education institutions need to have a shift in their culture, technological infrastructure, and teaching practices. In particular, a transformation from assessment for accountability to assessment for learning is necessary. It is recognised, however, that educational institutions lack the practical, technical, and financial capacity to effectively implement big data collection and learning analytics.

Network Analysis

Eight articles, see table 9, examined the use of network analysis in various learning environments. Many of the studies looked at Social Network Analysis to analyse or predict aspects of learner behaviour or learning outcomes. In particular, they studied informal networks in professional development, online community practices, personal learning environments, virtual learning environments, to understand or predict performance, understand network structures, interaction, or the influence of services.

de Laat & Schreurs (2013) researched the potential of learning analytics, in particular, network analysis, to *raise awareness of informal network activities in professional development*. The professional development of school teachers from 70 schools during an internship program for a school district was studied through their networks. They developed a learning analytics based tool, NAT, that analyses group discussions and presents visualisations of the work-related problems that the teaching professionals were working on in their various social networks. They argued that LA can shed light on on-going professional development activities connected with work practices, and LA can support a bottom-up culture of learning driven by the needs of professionals. The study also concluded that developing visualisations entails a significant amount of work. Moreover, bringing together separated networks is time-consuming, and school leadership does not recognise the importance of informal networking.

Johri & Teo (2013) examined the use and viability of learning analytics in informal learning. The *network structure of an online community of practice* is analysed through over 200 000 messages generated over ten years. Two methods were employed: motif analysis to discover the *patterns of community sub-structures*, and temporal analysis to determine the *effect of an external event on newcomer participation*. The study drew the conclusion that social network analysis methods are useful in community analysis.

Casquero et al. (2014) used Social Network Analysis (SNA) to analyse the *influence of an increased number of services in personal learning environments* on the connectivity and interaction strength of student's personal networks. It was discovered that offering more services does not significantly increase the number of ties among the students, but strengthens the interactions within a group.

Orduna et al. (2014) centred their research around data sets generated by students working in mobile remote laboratories. A sociocentric network perspective was used to *examine user interactions* on a network level. Similarity among the submitted exercises by students was determined in order to analyse the behaviour of students. It was found that students with high outdegree were sharing exercises with other students, while high indegree indicated that students were waiting for others to solve the tasks.

Table 9 Network Analysis articles

	Citation	Contribution Type	Data Client	Educational Setting	Data	Method	Data Subject	Pedagogical Approach	Learning Environment
1	Casquero et al. (2014)	Application Model Development Tool Development	Institutions	University	Log Data	Cluster Analysis Network Analysis	Learners	Collaborative Learning / CSCL Lifelong Learning	Online Learning
2	Casquero et al. (2016)	Application Model Development	Institutions	University	Log Data Performance Data	ANOVA Cluster Analysis Network Analysis RegressionAnalysis	Learners	Collaborative Learning / CSCL Lifelong Learning Social Learning	Online Learning
3	de Laat & Schreurs (2013)	Application Tool Development	Institutions	Informal Learning Workplace	Log Data Survey Data Text Data	Network Analysis Text Mining	Learners	Lifelong Learning Networked Learning	Online Learning
4	Hernandez-Garcia et al. (2015)	Application Model Development	Educators	University	Log Data Performance Data	ANOVA Correlation Analysis Data Visualization Descriptive Statistics Network Analysis	Educators Learners	Collaborative Learning / CSCL Group Learning Social Learning	Online Learning
5	Hernandez-Nanclares et al. (2016)	Application Model Development	Educators	University	Survey Data	Network Analysis Regression Analysis	Learners	Collaborative Learning / CSCL Group Learning	Blended Learning
6	Johri & Teo (2013)	Application Model Development	Insitutions	Informal Learning Workplace	Log Data	Descriptive Statistics Network Analysis	Learners	Interest-driven Learning	Online Learning
7	Mouri et al. (2015)	Application Tool Development	Learners	University	Log Data Pre- and Posttest Survey Data	Descriptive Statistics Network Analysis	Learners	Collaborative Learning / CSCL	Immersive Learning Environment Mobile Learning
8	Orduna et al. (2014)	Application	Educators	University	Log Data Text Data	Correlation Analysis Descriptive Statistics Network Analysis	Learners	Social Learning	Immersive Learning Environment Online Learning

Table 10 Overview articles

	Citation	Contribution Type	Data Client	Educational Setting	Data	Method	Data Subject	Pedagogical Approach	Learning Environment
1	Siemens (2013)	Model Development Theory	Researchers	-	-	-	-	-	-

Hernandez-Garcia et al. (2015) focused on *social learning analytics as a means to predict academic performance*. In particular, the impact of Social Network Analysis (SNA) visualisations on the learning process, as well as the relation between social network parameters and individual student, and class outcomes was studied. An analysis of the interactions in a virtual campus environment between students, consultant teachers, and the coordinating professor revealed that there is no evidence for a relation between social network parameters and class performance, while the amount of teacher activity is significant in influencing the numbers of drop-outs and collaborative learning processes.

Mouri et al. (2015) developed a mobile learning tool that uses network graphs and temporal analysis to support student language learning. Networks were built in relationship to words, places, and time. Moreover, students had access to the learning log dashboard, where they could get information on their *language progress*. Pre-test, post-test and surveys were used to evaluate the data logs results.

Casquero (2016) compared student use of virtual learning environments and personal learning environments, specifically looking at which environment helps students develop a larger social network and achieve a better learning performance. In particular, the *effect of the size of the personal network on the student learning performance* is studied. The study concluded that personal learning environments support larger, more distributed, and balanced personal network with research social capital than virtual learning environments. No significant difference in learning performance was found between the two environments, however, the larger size of a personal network helped students achieve a higher grade.

Hernandez-Nanclares et al. (2016) focused on *cross-boundary knowledge sharing for inter- and intra-group learning*. They demonstrated that Social Network Analysis (SNA) can support a more nuanced discussion of the merits and drawbacks of students crossing the group boundaries, and suggest that such analysis can help researchers to "better determine the optimal design for designing tasks that encourage strong inter-group learning links, strong intra-group learning links, or a combination of the the two." (p. 13).

Overview

The corpus contains one overview article, see table 10.

Siemens (2013) presented a detailed overview of the early (2013) emergence of learning analytics as a new discipline. It addresses, among other things, the most important definitions, the fields that have contributed technology and methodologies to the development of learning analytics, analytics models, the importance of learning analytics, and the need to increase the capacity for learning analytics application. In addition, the dimensions of implementing learning analytics/educational data mining in an institution are addressed in a learning

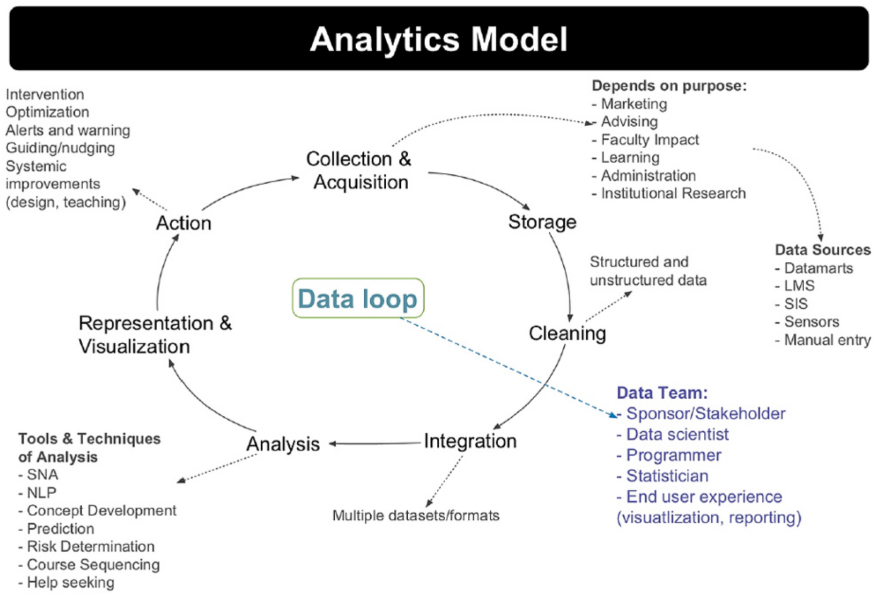


Figure 9 Learning analytics model (from Siemens, 2013, p. 1392)

analytics model, see figure 9. The LA model identifies the stakeholders, data sources, data team, and cycle of stages involved in carrying our learning analytics. What is interesting about the model is that it seems like pedagogical knowledge is missing from the data team. Siemens concludes that "managing large quantities of learner-generated data and gaining insight into the learning process through LA raise the profile of new tools and new techniques." (p. 1396), and this will require the development of academic programs to provide researchers and practitioners with the necessary skills.

Predictive Analysis

This theme has twenty articles that carried out predictive analysis to try to predict dropouts, dropout rates, academic performance, student success. Analytics methods such as correlation analysis, descriptive analysis, regression analysis, text analysis were used. Different metrics were identified and used, including one model that uses 200 variables. One observation is that all the articles addressed higher education or MOOCs, and a second observation is that predictive analysis was very context dependent and it is questionable whether a particular model could be transferred to another context within the same institution, let alone another institution. Table 11 presents the papers.

The theoretical background of the Gasevic et al. (2013) study is the theory of social capital that states that social interactions stimulate social and economic benefits. The first hypothesis proposed that student social capital accumulated in their courses correlated with their academic performance, while the second hypothesis predicted that cross-class networks and their influence on academic performance also correlated with academic performance. Using social network analysis predictors, as well as regression analysis, the study concluded that *both student social capital and their cross-class networks correlate positively with students' academic performance*.

Kotsiantis et al. (2013) studied the correlation between learning performance and interaction data extracted from Moodle. In addition to determining which data items are the best performance predictors, the main finding of this study was that *student perception towards Moodle and computer possession are the strongest predictors of student success*. Furthermore, there is a positive association between academic performance and Moodle usage.

Ognjanovic et al. (2016) developed a model that *predicts which courses students will choose in a specific academic term and year in the future*. They focused on which data sources from institutional student information systems are relevant for the *course prediction model*. To determine course selection course factors, such as course and instructor characteristics, course grades, and individual factors such as course timing, demographic data, and student interests, were taken into consideration. The results from the application of the model on an undergraduate degree program demonstrated that the accuracy of the student course predictions was high.

Agudo-Peregrina et al. (2014) examined if it is possible to determine system independent characteristics of learning interaction. The relationship between the interaction data and academic performance was based on log data analysis to determine if course characteristics such as instruction mode influence the results. Interactions based on the frequency of use, as well as participation mode were investigated. The study defined three characterisations of learning interaction. Moreover, *academic performance was correlated with interaction data in online courses with student-teacher interactions as a main predictor of student success*, while passive interactions and student-content interactions were not significant.

Table 11 Predictive Analysis articles

	Citation	Contribution Type	Data Client	Educational Setting	Data	Method	Data Subject	Pedagogical Approach	Learning Environment
1	Agudo-Peregrina et al. (2014)	Application Model Development	Educators Institutions	Informal Learning University Workplace	Log Data Performance Data	Correlation Analysis Descriptive Statistics Regression Analysis	Learners	Lifelong Learning	Blended Learning Online Learning
2	Akhtar et al. (2015)	Application Model Development	Educators Intitutions	University	Focus Group Data Log Data Observation Data Performance Data	ANOVA Correlation Analysis Regression Analysis	Learners	Collaborative Learning / CSCL Group Learning	Online Learning
3	Bilkstein et al. (2014)	Application Model Development	Educators	University	Log Data Performance Data Text Data	Cluster Analysis Correlation Analysis Hidden Markov Model Regression Analysis	Learners	Inquiry-based Learning Project-based Learning	-
4	Calvert (2014)	Application Model Development	Insitutions	University	Demographic Data Learning Design Data Log Data Performance Data	Regression Analysis	Learners	-	Online Learning
5	Cambruzzi et al. (2015)	Application Model Development Tool Development	Insitutions	University	Log Data Performance Data Text Data	Artificial Neural Networks Data Visualization Natural Learning Processing Network Analysis	Learners	-	Online Learning
6	de Barba et al. (2016)	Application Model Development	Institutions	MOOCs	Log Data Performance Data Survey Data	Correlation Analysis Descriptive Statistics Structural Equation Modeling	Learners	Mastery Learning Self-directed / Self-regulated Learning	Online Learning
7	Fidalgo-Blanco et al. (2015)	Application Model Development	Educators	University	Log Data Performance Data	Correlation Analysis Descriptive Statistics	Learners	Collaborative Learning / CSCL Group Learning	Blended Learning
8	Gasevic et al. (2013)	Application Model Development	Institutions	University	Log Data Performance Data	Descriptive Statistics Network Analysis Regression Analysis	Learners	Group Learning Networked Learning Social Learning	Online Learning
9	Ifenthaler & Widanapathirana (2014)	Application Model Development	Institutions	University	Demographics Data Log Data Performance Data	Regression Analysis Support Vector Machines	Learners	-	Online Learning

Table 11 Predictive Analysis articles (continued)

	Citation	Contribution Type	Data Client	Educational Setting	Data	Method	Data Subject	Pedagogical Approach	Learning Environment
10	Jo et al. (2015)	Application Model Development	Educators	Workplace	Log Data Performance Data	Correlation Analysis Descriptive Statistics Regression Analysis	Learners	Self-directed / Self-regulated Learning	Online Learning
11	Joksimovic et al. (2015b)	Application Model Development	Educators	University	Log Data Performance Data Text Data	Correlation Analysis Descriptive Statistics Natural Learning Processing Regression Analysis	Learners	-	Online Learning
12	Kotsiantis et al. (2013)	Application Model Development	Institutions	University	Log Data Performance Data	Association Rules Cluster Analysis Data Visualization Decision Tree Learning	Learners	Constructionist / Constructivist Learning Collaborative Learning / CSCL Social Learning	Blended Learning
13	Lonn et al. (2015)	Application Model Development	Educators	University	Demographic Data Log Data Performance Data Survey Data	Data Visualization Regression Analysis T-Test	Learners	Mastery Learning	Blended Learning
14	Ognjanovic et al. (2016)	Application Model Development	Institutions	University	Demographic Data Learning Design Data Log Data	ANOVA Correlation Analysis Descriptive Statistics Neural Networks T-Test	Learners	-	Online Learning
15	Strang (2016)	Application Model Development	Educators Institutions	University	Demographic Data Log Data Performance Data Pre- and Posttest Text Data	ANOVA Cluster Analysis Correlation Analysis Descriptive Analysis Natural Language Processing Regression Analysis T-Test	Learners	Reflective Learning	Online Learning
16	Tempelaar et al. (2015)	Application Model Development	Educators	University	Demographic Data Log Data Performance Data	Correlation Analysis Regression Analysis	Learners	Adaptive / Personalized Learning Group Learning Problem-based Learning	Blended Learning
17	Vu et al. (2015)	Application Model Development	Educators	MOOCs	Log Data	Descriptive Statistics Network Analysis Regression Analysis	Learners	Collaborative Learning / CSCL Networked Learning Peer-based Learning Social Learning	Online Learning
18	Xing et al. (2016	Application Model Development	Educators Institutions	MOOCs	Log Data	Bayesian Network Decision Tree Learning Descriptive Statistics Genetic Programming Network Analysis Principle Component Analysis Regression Analysis	Learners	Social Learning	Online Learning
19	You (2016)	Application Model Development	Educators	University	Demographic Data Log Data Performance Data	Correlation Analysis Descriptive Statistics Regression Analysis	Learners	Self-directed / Self-regulated Learning	Online Learning
20	Zacharis (2015)	Application Model Development	Educators	University	Log Data Performance Data	ANOVA Correlation Analysis Descriptive Statistics Regression Analysis T-Test	Learners	Collaborative Learning / CSCL Self-directed / Self-regulated Learning	Blended Learning

Bilkstein et al. (2014) performed a comprehensive analysis of the programming learning behaviour patterns of university students. A variety of methods such as clustering and regression analysis were used in three experiments. Not only was a temporal analysis of programming patterns performed but in addition *students behaviour patterns were correlated with their academic performance*.

Calvert (2014) used predictive modelling to analyse student retention at important milestones in modules at the Open University, UK. Around 200 variables were used to build the prediction model and a number of *interventions were designed to support students*, as well as evaluate the current curriculum.

Ifenthaler & Widanapathirana (2014) developed a *framework that uses Support Vector Machines to evaluate predictive models* that use a variety of data such as student demographic data, academic records, and even parent's educational background.

Cambruzzi et al. (2015) described the data that can be collected in university distance education courses and how this data can help with retention. A case study based on the collected data was carried out to *predict dropout rates*. Afterwards, a set of interventions were applied and the dropout rate was reduced.

Fidalgo-Blanco et al. (2015) applied learning analytics techniques in order to improve assessment of teamwork, with a focus on individual assessment of each group member. The *number of messages created and viewed by a student were correlated with the individual grade*s and it was found that the *biggest predictor of student success is active student-student interactions*.

Jo et al. (2015) analysed time management strategies in an online course to determine performance predictors. Using correlation analysis and multiple linear regression, it was found that *regular login intervals are a strong predictor of a student success*.

Joksimovic et al. (2015b) used descriptive statistics, multiple regression, and text analysis to analyse the interactions in a discussion forum in order to determine the *relationship between social presence in a forum, academic performance, and changes in the instructional design*. One result is *course design that increased the level of meaningful interactions between students had a significant impact on the development of social presence*, and thus could *positively affect students' academic performance*, and a second that *indicators of social presence can be used for early detection of students at risk of failing a course*.

Lonn et al. (2015) examined the changes in student motivation over time, as well as the level of their mastery orientation (i.e., their competence and task mastery). The analysis of the interaction data from an LMS system, survey results, and demographic variables showed that *the mastery orientation decreases over time, when students are exposed to the visualization of their performance data*. This negative result leads them to conclude that "student perceptions of their goals and formative performance need to be carefully considered in the design of learning analytics interventions since the resulting tools can affect students' interpretations of their own data as well as their subsequent academic success" (p. 90).

Tempelaar et al. (2015) extracted and analysed LMS data from a mathematics course to determine the relationship between academic performance and student learning dispositions. The data was analysed using multiple regression. The results show that *LMS click data is not a good predictor of student success*, while *data from formative assessment activities has the strongest predictive power of academic performance*.

Vu et al. (2015) proposed three extensions to the relational event framework to *model the co-evolution of multiple network event streams from online applications*. Three modelling problems were addressed: "(1) the utility of social factors, performance indicators, and clickstream behaviours in the prediction of course dropout, (2) the social and temporal structure of learner interactions across discussion threads, and (3) the forms of mutual dependence of social learning interactions on prior learning success and future learning success on forms of prior social learning interaction." (p. 121).

Zacharis (2015) used correlation and regression analysis to find a link between interaction data and academic performance, as well as determining which variables are the best at predicting student success. The study found that r*eading and posting message in the discussion forum were correlated positively with the final grades*. Also, the *engagement with online quizzes influences positively student success*.

de Barba et al. (2016) investigated how motivation and participation influence student performance in a MOOC, and they propose a *structural equation model to predict student performance in a MOOC*. The two main *variables of performance are interaction data logs and the state of motivation* measured by a survey.

Xing et al. (2016) examined the dropout rates in a MOOC in order to test predictive and temporal models that prioritises at-risk students to *determine their likelihood to drop out of a course*. The study found that an ensemble stacking prediction model, which supports more robust and accurate prediction models, had the best performance.

You (2016) identified the variables that can predict student performance in a LMS, including *regular study, late submission of assignments, number of sessions (course logins), and proof of reading the course information packets significantly predicted their course achievement*. Moreover, the identified indicators and academic records from the middle of the course were analysed to determine if they could predict student final scores and the results show that these indicators collected in the middle of the course significantly predicted course achievement.

Akhtar et al. (2015) analysed interaction data from a CSCL learning environment using ANOVA and Pearson correlation. Furthermore, additional data was gathered during focus group interviews. One of the main findings was that there was a *positive correlation between performance and attendance, and time spent on task*.

Strang (2016) tested 10 hypotheses regarding the influence of demographic and interaction variables on student performance. Descriptive statistics, correlation analysis, multiple regression analysis, and ANOVA are some of the methods used. The study found that only *course logins are significantly related to course outcomes*.

Text Analysis

In this theme there are three articles that use learning analytics to analyse texts, see table 12, such as essays and discussion forum messages.

Lei et al. (2014) used text analysis techniques to *analyse and automatically grade student essays*. Not only metadata such as number of words is collected and analysed, but also semantic aspects such as verb cohesion and argument overlap is collected.

Tobarra et al. (2014) applied network analysis and semantic analysis to *analyse discussion forum messages*. The study tried to detect both student behaviour patterns and topics of discussion. Two algorithms were developed to facilitate an automatic topic detection.

Table 12 Text Analysis articles

	Citation	Contribution Type	Data Client	Educational Setting	Data	Method	Data Subject	Pedagogical Approach	Learning Environment
1	Al-Shmoery et al. (2015)	Tool Development	Researchers	-	-	-	-	-	-
2	Lei et al. (2014)	Application Tool Development	Learners	University	Text Data	Bayesian Network Descriptive Statistics	Learners	Self-directed / Self-regulated Learning	-
3	Tobarra et al. (2014)	Application Model Development	Institutions	University	Log Data Text Data	Correlation Analysis Descriptive Statistics Natural Learning Processing Network Analysis	Learners	Collaborative Learning / CSCL	Online Learning

Table 13 Tool Development articles

	Citation	Contribution Type	Data Client	Educational Setting	Data	Method	Data Subject	Pedagogical Approach	Learning Environment
1	Bull & Kay (2016)	Model Development Tool Development	Intitutions	-	-	-	-	-	-
2	Kickmeier-Rust et al. (2014)	Application Tool Development	Educators Learners	Primary School	Demographic Data Log Data Survey Data	Descriptive Statistics	Learners	Game-based Learning	Immersive Learning Environment
3	Pardos et al. (2016)	Model Development Tool Development	Institutions	-	-	-	-	-	-
4	Tabuenca et al. (2015)	Application Tool Development	Learners	University	Log Data Performance Data Survey Data	ANOVA Correlation Analysis Data Visualization Descriptive Statistics	Learners	Lifelong Learning Self-directed / Self-regulated Learning	Mobile Learning Online Learning

Al-Shmoery et al. (2015) introduced a new tool that can be used for text analysis using, among others, *semantic similarity on discussion forum messages*. Moreover, the article presents other available tools for text analysis and compares their features with the newly developed tool.

Tool Development

Four articles, see table 13, presented tools they have developed that use learning analytics to assist with formative feedback, self-regulated learning, visualisation of learner models, and course model, participant characteristics and behaviour visualisations for MOOCs.

Kickmeier-Rust et al. (2014) developed *an app that helps primary school students learn mathematics*. It examines the link between live formative feedback and learning performance, and analyses the results by gender and overall by applying statistical usage methods and surveys.

Tabuenca et al. (2015) explored the *impact of learning analytics on self-regulated learning in a mobile learning environment*. Logging time patterns were explored. Also, timing of notifications was analysed. There was no correlation found between the final grades and the time-logs.

Bull & Kay (2016) presented *SMILI, a tool to facilitate open learner models*. This study described the tool's features, the underlying framework, and visualisations produced.

Pardos et al. (2016) developed *a learning analytics architecture, including a web-based data repository and analysis tool, for MOOCs*, moocRP, that can be used to distribute, analyse and visualise data. The study described the tool in the contexts of interoperability, security, data mining, data distribution, and data models. The authors argued that such an open learning analytics platform is required in order to move the adoption of learning analytics forward.

Visualisations

Seven articles examined the use of learning analytics in helping visualisation development. A number of tools were developed to use SNA on discussion forums, improve student participation in collaborative tasks, etc. Research also looked at learning dashboards, and if teachers and students understand visualisations. Table 14 presents the visualisation articles.

McCormick (2013) developed a tool, *SNAPP, that uses social network analysis on discussion forum data*. The tool supports teachers by visualising the online discussions. Three instructors tested the tool and were interviewed to share their opinions about SNAPP.

Verbert et al. (2014) is an overview paper about l*earning dashboards*. Not only are dashboard tools presented but also the underlying theories on dashboards and various aspects of dashboard implementation.

Melero et al. (2015) carried out research on *how visualisations can support inquiry and self-assessment in location-based learning*. Both students as well as teachers were involved in the case studies. The study concluded that visualisations of log data is helpful not only for teachers to redesign and develop inquiry activities, but also for students to better assess their performance.

Papamitsiou & Economides (2015) experimented with *Temporal Learning Analytics Visualizations* (TLAVs) to examine if a student's temporal data can explain their behaviour during assessment and if teachers can interpret the visualisations The tool was evaluated by secondary school teachers through two iterations and both usability and usefulness of the visualisations were assessed, and the results showed promise in helping teacher awareness of the class and individual student progress.

Table 14 Visualisations articles

	Citation	Contribution Type	Data Client	Educational Setting	Data	Method	Data Subject	Pedagogical Approach	Learning Environment
1	Kim et al. (2016)	Application Tool Development	Learners	University	Log Data Survey Data	Correlation Analysis Descriptive Statistics Regression Analysis T-Test	Learners	Self-directed / Self-regulated Learning	Online Learning
2	McCormick (2013)	Tool Development	Educators	University	Survey Data Text Data	Data Visualization Network Analysis	Educators	-	Online Learning
3	Melero et al. (2015)	Application	Educators Learners	High School	Focus Group Data Log Data Pre- and Posttest Survey Data	Correlation Analysis Data Visualization Descriptive Statistics T-Test	Educators Learners	Active Learning	Immersive Learning Environment Mobile Learning
4	Papamitsiou & Economides (2015)	Model Development Tool Development	Educators Learners	-	Focus Group Data Survey Data	Correlation Analysis Data Visualization Descriptive Statistics Regression Analysis Structural Equation Modeling	Educators	-	Online Learning
5	Ruipérez-Valiente et al. (2015)	Model Development Tool Development	Educators Learners	University	Log Data	Data Visualization Descriptive Statistics	Learners	-	Online Learning
6	Verbert et al. (2014)	Theory	Researchers	-	-	-	-	-	-
7	Zorilla et al. (2015)	Application Model Development	Educators	University	Log Data Observation Data	Cluster Analysis Correlation Analysis Data Visualization	Educators Learners	Collaborative Learning / CSCL	Online Learning

Ruiperez-Valiente et al. (2015) introduced the ALAS-KA add-on to the Khan Academy platform that is used to *visualise students' interaction data individually and as a group*. The goal was to provide additional information about the students to help teachers and students make decisions in the learning process.

Zorilla et al. (2015) developed a customisable formulae that allowsed *instructors to set up particular parameters or indicators* to assess activity performance by students such as student attendance in a LMS, and examined if visualisations of these indicators can help to better understand the learning process, detect dropout risks, and if they correlate with student performance.

Kim et al. (2016) examined the *effectiveness of dashboards*, in particular if there is a relationship among the frequency of dashboard usage, student satisfaction, and learning performance. Using log data analysis as well as pre- and post-test self-reports it could be concluded that access to the dashboard increases learning achievement, while less dashboard usage increased the dashboard satisfaction. Furthermore, frequent usage and high academic performance correlated with a lower dashboard satisfaction.

3.2 Data & Methods

In order to answer research question 2

What key data and methods are being used?

we identified the data and analytics methods being used in the studies described in each article. In this section we present these results.

Data

There are a number of data types that are used for learning analytics in the articles, see figure 10. *Log data* is generated by the use of a tool, *observation data* is written down while sitting in the classroom, *learning design data* includes data on the content of courses, curriculum, the instructional design, *text data* is text generated by learning activity such as discussion forum participation or essay writing, *performance data* is final grade data, exam scores, etc., *survey data* is the results from a questionnaire or survey, *demographic data* includes background demographics of a leaner, *focus group data* is recorded during a group interview, *pre- and post-test data* is gathered from tests given to study participants.

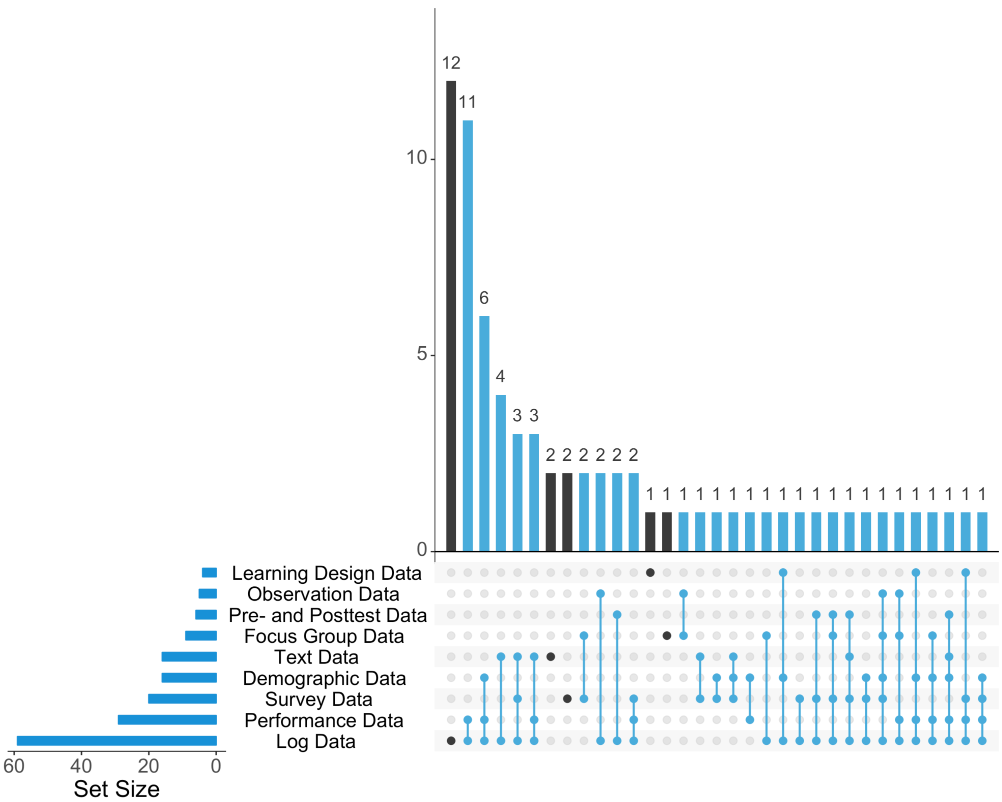


Figure 10 Sets and intersections of the data types

The most common *single dataset* is log data (12 articles), and other single datasets used include text data (3 articles), learning design data and focus group data (both used in 1 article). Several studies use *two data sets*, the most common paring being log data and performance data, while others use log data and text data (4), survey data and focus group data (2), log data and pre- and post-test data (2), log data and observation data (2), focus group data and observation data (1), and log data and focus group data (1). Other studies used *three data sets*, with the most popular combination being log data, performance data, and demographic data (6), and other examples being log data, survey data and text data (3), log data, performance data, and survey data (2). Six studies use 4 datasets, including log data, survey data, focus group data, and pre- and post-test data (1), and another using log data, performance data, demographic data, and learning design data (1). Two studies used 5 datasets, including one study using log data, survey data, demographic data, focus group data, and observation data (1). Table 16 lists the data types by article, table 15 summarises the number of articles that use 1-5 data sources, and figure 11 visualises this.

Table 15 Number of data sources by article

Number of Data Sources	Freq	%
1	18	25%
2	27	37%
3	18	25%
4	7	10%
5	2	3%
Total	72	

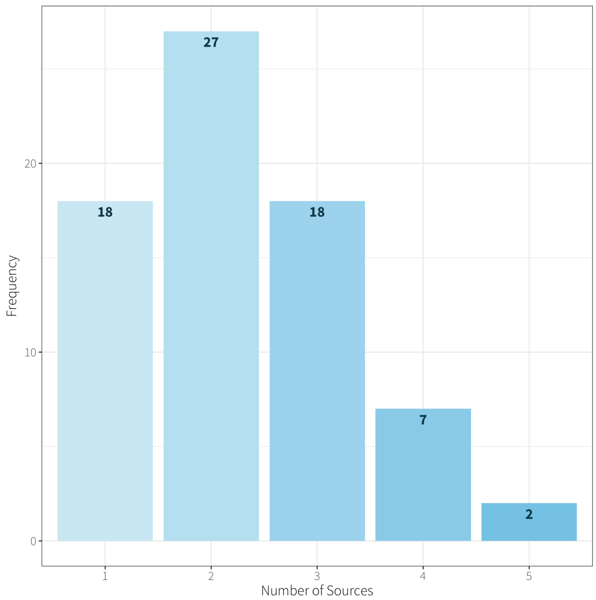


Figure 11 Number of data sources by articles

Table 16 Data types by article

Log Data	Performance Data	Survey Data	Text Data	Demographic Data	Pre- and Posttest	Focus Group/Interviews	Learning Design Data	Observation Data	Citation
✓									Berland et al. (2013), Casquero et al. (2014), Goggins et al. (2015), Johri & Teo (2013), Kennedy et al. (2013), Leony et al. (2015), Ma et al. (2015), Ruiperez-Valiente et al. (2015), Serrano-Laguna et al. (2014), Vu et al. (2015), Xing et al. (2015), Xing et al. (2016)
		✓							Hernandez-Nanclares et al. (2016), Vahdat et al. (2016)
			✓						Berland et al. (2015), Lei et al. (2014)
						✓			McKenney & Mor (2015)
							✓		Toetenel & Rienties (2016)
✓	✓								Agudo-Peregrina et al. (2014), Brooks et al. (2014), Casquero et al. (2016), del Puerto Paule-Ruiz et al. (2015), Fidalgo-Blanco et al. (2015), Gasevic et al. (2013), Hernandez-Garcia et al. (2015), Jo et al. (2015), Joksimovic et al. (2015a), Kotsiantis et al. (2013), Zacharis (2015)
✓		✓							Kim et al. (2016)
✓			✓						Haya et al. (2015), Orduna et al. (2014), Tobarra et al. (2014), Xie et al. (2014b)
✓					✓				Halverson & Owen (2014), Martin et al. (2015)
✓						✓			Rodriguez-Triana et al. (2013)
✓								✓	Zorilla et al. (2015), Thompson et al. (2013)
	✓			✓					Yasmin (2013)
		✓	✓						McCormick (2013)
		✓		✓					Florian-Gaviria et al. (2013)
		✓				✓			Munoz-Merino et al. (2015), Papamitsiou & Economides (2015)
						✓		✓	Rodriguez-Triana et al. (2015)
✓	✓	✓							de Barba et al. (2016), Tabuenca et al. (2015)
✓	✓		✓						Bilkstein et al. (2014), Cambruzzi et al. (2015), Joksimovic et al. (2015b)
✓	✓			✓					de Freitas et al. (2015), Ifenthaler & Widanapathirana (2014), Liu et al. (2016), Tempelaar et al. (2015), Volk et al. (2015), You (2016)
✓		✓	✓						de Laat & Schreurs (2013), Van Leeuwen et al. (2014), Van Leeuwen et al. (2015)
✓		✓		✓					Kickmeier-Rust et al. (2014)
✓		✓			✓				Mouri et al. (2015)
✓				✓			✓		Ognjanovic et al. (2016)
		✓	✓	✓					Nistor et al. (2014)
✓	✓	✓		✓					Lonn et al. (2015)
✓	✓	✓					✓		Rienties & Toetenel (2016)
✓	✓			✓		✓			Gibson & de Freitas (2016)
✓	✓			✓			✓		Calvert (2014)
✓	✓					✓		✓	Akhtar et al. (2015)
✓		✓	✓		✓				Yen et al. (2015)
✓		✓			✓	✓			Melero et al. (2015)
✓	✓		✓	✓	✓				Strang (2016)
✓		✓		✓		✓		✓	Xie et al. (2014a)

Methods

Figure 12 shows the frequency of the *data analysis methods* used in the corpus. By far the most commonly used is *descriptive statistics*, which was reported in 43 articles (43%). The second most used was *correlation analysis*, 36, followed by *regression analysis,* 24, *network analysis,* 16, *cluster analysis and data visualisations*, each 13, *ANOVA and T-Test*, each 10. The remainder of the methods were reported in 1-5 articles. Some of these less used approaches are the artificial intelligence methods such as *genetic programming*, *bayesian networks, neural networks*, while *multimodal analysis* uses a combination of data such as video, audio, text, speech, and combines various methods such as computer vision, text analysis, etc. to tell the story in the data.

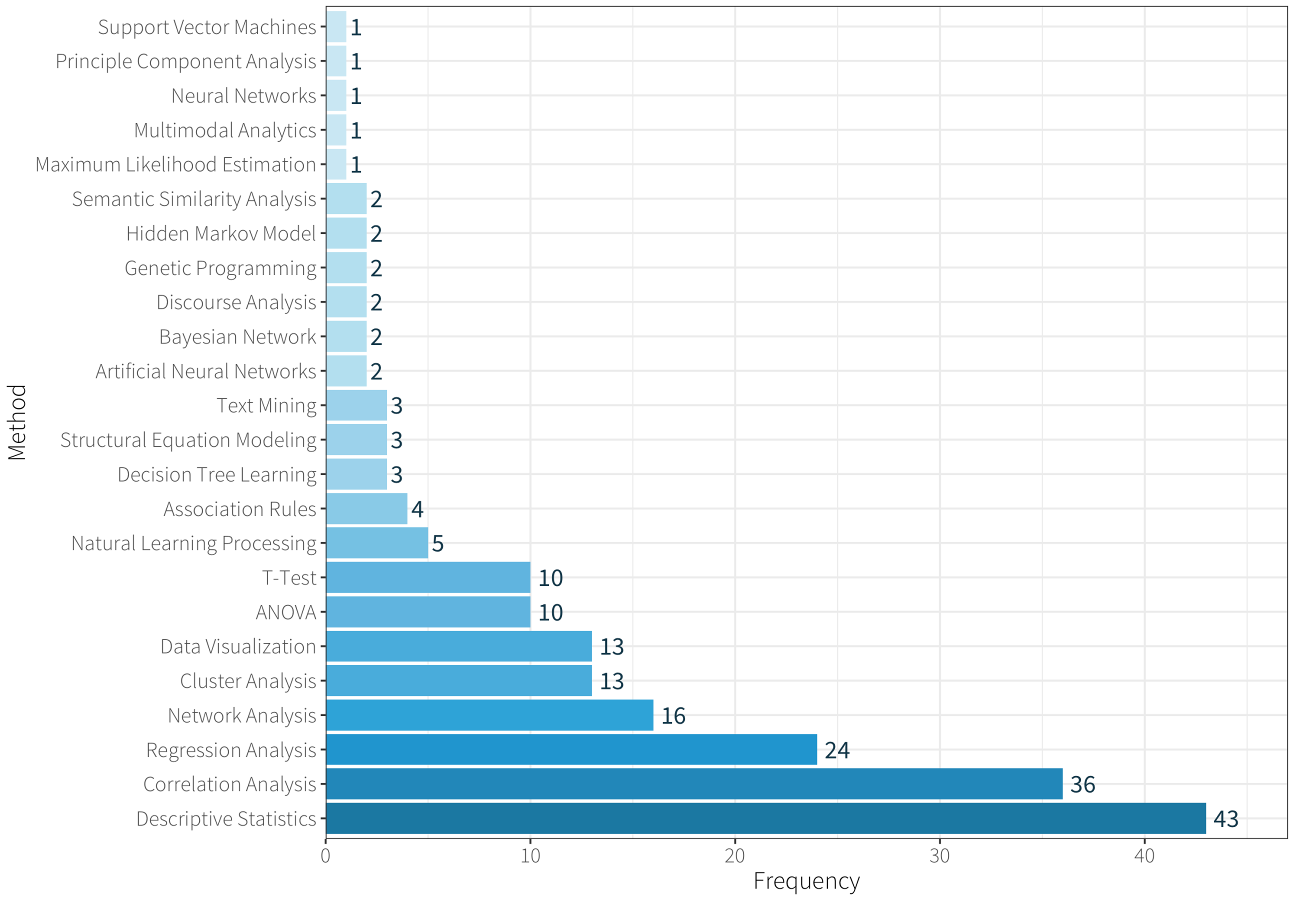


Figure 12 Method frequency

Table 17 shows the methods used per article. There are only 7 articles that report using only 1 method.

Table 17 Methods by article

Citation	ANOVA	Artificial Neural Networks	Association Rules	Bayesian Network	Cluster Analysis	Correlation Analysis	Data Visualization	Decision Tree Learning	Descriptive Statistics	Discourse Analysis	Genetic Programming	Hidden Markov Model	Maximum Likelihood Estimation	Multimodal Analytics	Natural Learning Processing	NetworkAnalysis	Neural Networks	Principle Component Analysis	Regression Analysis	Semantic Similarity Analysis	Structural Equation Modeling	Support Vector Machines	T-Test	Text Mining
del Puerto Paule-Ruiz et al. (2015)			✓																					
Berland et al. (2013), de Freitas et al. (2015)					✓																			
Serrano-Laguna et al. (2014)							✓																	
Kickmeier-Rust et al. (2014)									✓															
Xing et al. (2015)											✓													
Calvert (2014)																			✓					
Ma et al. (2015)																					✓			
Kennedy et al. (2013)			✓									✓												
Lei et al. (2014)				✓					✓															
Brooks et al. (2014), Goggins et al. (2015)					✓	✓																		
Casquero et al. (2014)					✓											✓								
Fidalgo-Blanco et al. (2015), Halverson & Owen (2014), Leony et al. (2015), Toetenel & Rienties (2016), Xie et al. (2014a)						✓			✓															
Tempelaar et al. (2015)						✓													✓					
Ruiperez-Valiente et al. (2015)							✓		✓															
McCormick (2013)							✓									✓								
Yasmin (2013)								✓	✓															
Johri & Teo (2013), Mouri et al. (2015)									✓							✓								
Berland et al. (2015)									✓											✓				
Hernandez-Nanclares et al. (2016)																✓			✓					
de Laat & Schreurs (2013)																✓								✓
Ifenthaler & Widanapathirana (2014)																			✓			✓		
Martin et al. (2015)	✓				✓				✓															
Akhtar et al. (2015)	✓					✓													✓					
Zorilla et al. (2015)					✓	✓	✓																	
Yen et al. (2015)						✓			✓	✓														
Orduna et al. (2014)						✓			✓							✓								
Agudo-Peregrina et al. (2014), Jo et al. (2015), Rienties & Toetenel (2016), Volk et al. (2015), You (2016)						✓			✓										✓					
de Barba et al. (2016)						✓			✓												✓			
Lonn et al. (2015)							✓												✓				✓	
Thompson et al. (2013)									✓	✓				✓										
Vu et al. (2015), Gasevic et al. (2013)									✓							✓			✓					
Van Leeuwen et al. (2015)									✓										✓				✓	
Van Leeuwen et al. (2014)									✓														✓	✓

Table 17 Methods by article (continued)

Citation	ANOVA	Artificial Neural Networks	Association Rules	Bayesian Network	Cluster Analysis	Correlation Analysis	Data Visualization	Decision Tree Learning	Descriptive Statistics	Discourse Analysis	Genetic Programming	Hidden Markov Model	Maximum Likelihood Estimation	Multimodal Analytics	Natural Learning Processing	Network Analysis	Neural Networks	Principle Component Analysis	Regression Analysis	Semantic Similarity Analysis	Structural Equation Modeling	Support Vector Machines	T-Test	Text Mining
Casquero et al. (2016)	✓				✓											✓			✓					
Tabuenca et al. (2015), Liu et al. (2016)	✓					✓	✓		✓															
Cambruzzi et al. (2015)		✓					✓								✓	✓								
Gibson & de Freitas (2016)		✓	✓		✓	✓																		
Kotsiantis et al. (2013)			✓		✓		✓	✓																
Bilkstein et al. (2014)					✓	✓						✓							✓					
Melero et al. (2015)						✓	✓		✓														✓	
Tobarra et al. (2014)						✓			✓						✓	✓								
Joksimovic et al. (2015b)						✓			✓						✓				✓					
Haya et al. (2015)						✓			✓							✓				✓				
Kim et al. (2016)						✓			✓										✓				✓	
Xie et al. (2014b)						✓			✓														✓	✓
Munoz-Merino et al. (2015)	✓				✓	✓	✓		✓															
Hernandez-Garcia et al. (2015)	✓					✓	✓		✓							✓								
Ognjanovic et al. (2016)	✓					✓			✓								✓						✓	
Zacharis (2015)	✓					✓			✓										✓				✓	
Papamitsiou & Economides (2015)						✓	✓		✓										✓		✓			
Joksimovic et al. (2015a)						✓			✓				✓						✓				✓	
Nistor et al. (2014)						✓			✓						✓	✓			✓					
Strang (2016)	✓				✓	✓			✓						✓				✓				✓	
Xing et al. (2016)				✓				✓	✓		✓					✓		✓	✓					

3.3 Characteristics of the Studies

In order to answer research question 3

What are the characteristics of the learning analytics studies?

we identified the type of contribution, the educational settings, data clients, data subjects, pedagogical approach, and learning environments in the studies reported in the articles. In this section we present these results.

Type of Contribution

Table 18 presents the four types of contributions that we have identified that an article can make: application, model, theoretical, or tool, or a combination. Figure 13 shows that the majority of articles apply learning analytics and develop a model (43 articles), while the next highest contribution is theory (16) followed by the application of learning analytics and tool development (10), and model and tool development (7), the application of learning analytics, model and tool development (7), tool development alone (4), theory and model development (4), and model development alone (3). Table 19 presents the types of contributions by article.

Table 18 Contribution types

Contribution type	Explanation
application	The article describes the application of learning analytics in an educational setting; can be descriptive, prescriptive, or predictive.
model development	The article describes the development of, or presents a model, framework or algorithm
theory	The article is theoretical.
tool development	The article describes the software development of a tool, or presents or evaluates a tool.

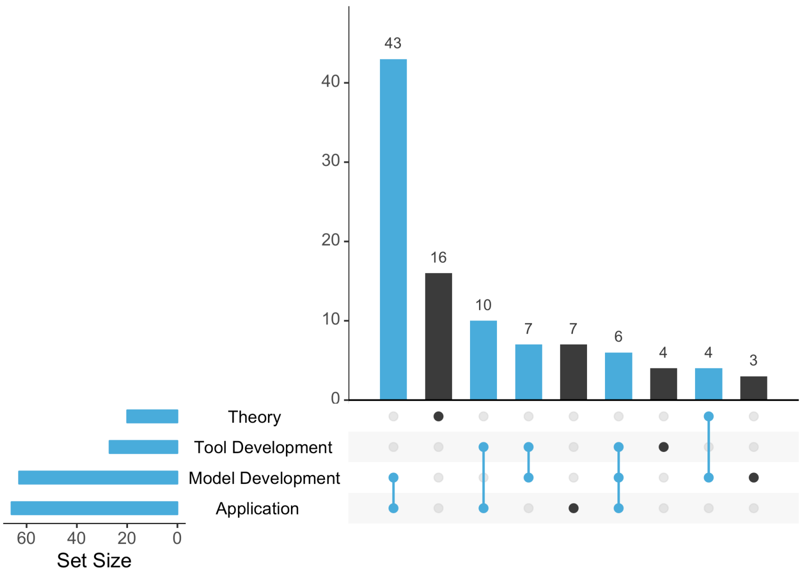


Figure 13 Sets and intersections of the contribution types

Table 19 Contribution types by article

Citation	Application	Model Development	Theory	Tool Development
Johri & Teo (2013), Liu et al. (2016), Martin et al. (2015), Melero et al. (2015), Orduna et al. (2014), Toetenel & Rienties (2016), Van Leeuwen et al. (2014)	✓			
Dix & Leavesley (2015), Drachsler & Kalz (2016), Macfayden et al. (2014)		✓		
Berland et al. (2014), Ellis (2013), Lockyer et al. (2013), Lundie (2015), Pardo & Siemens (2014), Persico & Pozzi (2015), Prinsloo & Slade (2014), Prinsloo et al. (2015), Robert-Mahoney et al. (2016), Rubel & Jones (2016), Slade & Prinsloo (2013), Thompson & Cook (2016), Verbert et al. (2014), Williams (2014), Williamson (2015), Williamson (2016)			✓	
Al-Shmoery et al. (2015), McCormick (2013), McKenney & Mor (2015), Santos et al. (2015)				✓
Agudo-Peregrina et al. (2014), Akhtar et al. (2015), Berland et al. (2013), Bilkstein et al. (2014), Brooks et al. (2014), Calvert (2014), Casquero et al. (2016), de Barba et al. (2016), de Freitas et al. (2015), del Puerto Paule-Ruiz et al. (2015), Fidalgo-Blanco et al. (2015), Gasevic et al. (2013), Gibson & de Freitas (2016), Hernandez-Garcia et al. (2015), Hernandez-Nanclares et al. (2016), Ifenthaler & Widanapathirana (2014), Jo et al. (2015), Joksimovic et al. (2015a), Joksimovic et al. (2015b), Kotsiantis et al. (2013), Leony et al. (2015), Lonn et al. (2015), Ma et al. (2015), Munoz-Merino et al. (2015), Nistor et al. (2014), Ognjanovic et al. (2016), Rienties & Toetenel (2016), Strang (2016), Tempelaar et al. (2015), Thompson et al. (2013), Tobarra et al. (2014), Vahdat et al. (2016), Volk et al. (2015), Vu et al. (2015), Xie et al. (2014a), Xie et al. (2014b), Xing et al. (2015), Xing et al. (2016), Yasmin (2013), Yen et al. (2015), You (2016), Zacharis (2015)	✓	✓		
Berland et al. (2015), de Laat & Schreurs (2013), Kennedy et al. (2013), Kickmeier-Rust et al. (2014), Kim et al. (2016), Lei et al. (2014), Mouri et al. (2015), Serrano-Laguna et al. (2014), Tabuenca et al. (2015), Van Leeuwen et al. (2015)	✓			✓
Dawson & Siemens (2014), de Laat & Prinsen (2014), Ferreira & Andrade (2014), Siemens (2013)		✓	✓	
Bull & Kay (2016), Florian-Gaviria et al. (2013), Fulantelli et al. (2015), Papamitsiou & Economides (2015), Pardos et al. (2016), Rodríguez-Triana et al. (2013), Ruiperez-Valiente et al. (2015)		✓		✓
Cambruzzi et al. (2015), Casquero et al. (2014), Goggins et al. (2015), Halverson & Owen (2014), Haya et al. (2015), Rodríguez-Triana et al. (2015)	✓	✓		✓

Educational Setting

Figure 14 shows the educational settings in which the studies took place (if it was possible to identify). The educational settings are divided into formal education (workplace, university, high school, middle school, primary school) and informal education. Research settings are excluded from this category (e.g., McKenney & Mor (2015), Papamitsiou & Economides (2015)). As 47 of the 63 studies where it was possible to identify the educational setting took place in universities it is clear that the majority of learning analytics studies take place in higher education. The 4 MOOC studies and the single workplace studies most likely include adults, so we can say that the typical learner in learning analytics is 18+ years of age. There are only 5 studies in higher school, 4 in middle school, and 2 in primary school. Table 20 shows educational settings per article.

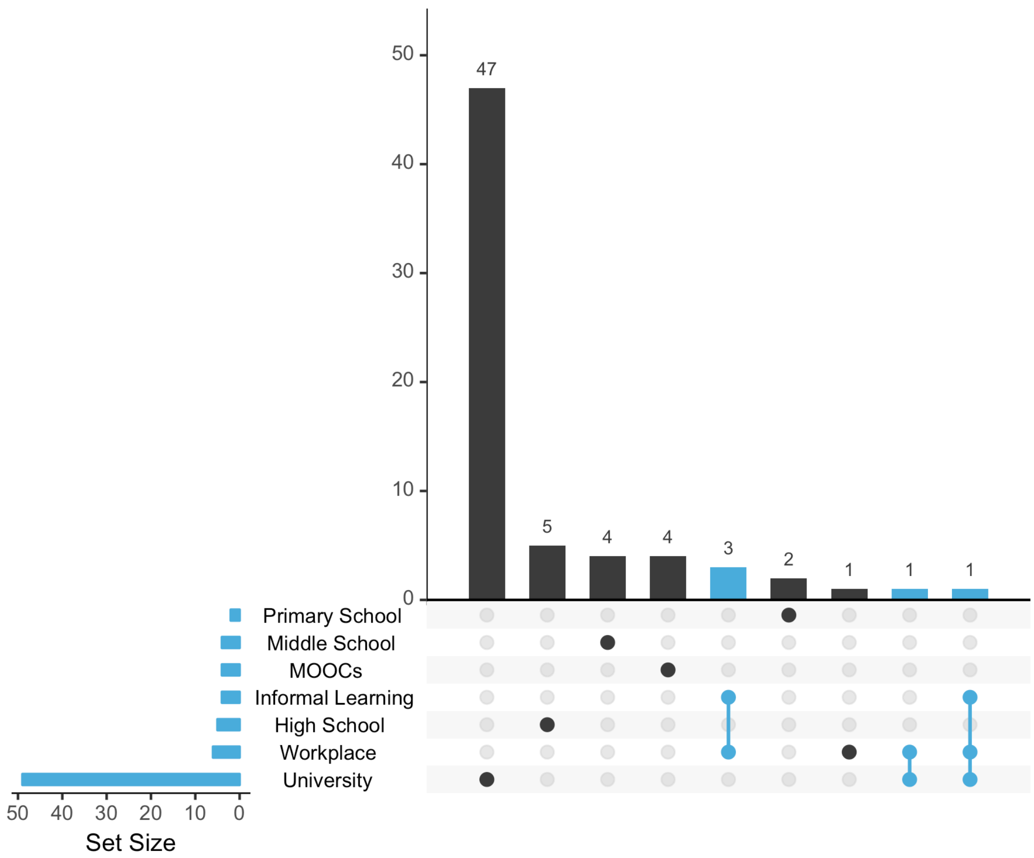


Figure 14 Sets and intersections of the educational settings

Table 20 Educational settings by article

Primary School	Middle School	High School	University	Workplace	Informal Learning	MOOCs	Citation
✓							Kickmeier-Rust et al. (2014), Martin et al. (2015)
	✓						Halverson & Owen (2014), Liu et al. (2016), Volk et al. (2015), Xing et al. (2015)
		✓					Berland et al. (2013), Berland et al. (2015), Melero et al. (2015), Xie et al. (2014a), Xie et al. (2014b)
			✓				Akhtar et al. (2015), Bilkstein et al. (2014), Brooks et al. (2014), Calvert (2014), Cambruzzi et al. (2015), Casquero et al. (2014), Casquero et al. (2016), de Freitas et al. (2015), del Puerto Paule-Ruiz et al. (2015), Fidalgo-Blanco et al. (2015), Gašević et al. (2013), Gibson & de Freitas (2016), Haya et al. (2015), Hernandez-Garcia et al. (2015), Hernandez-Nanclares et al. (2016), Ifenthaler & Widanapathirana (2014), Joksimovic et al. (2015a), Joksimovic et al. (2015b), Kennedy et al. (2013), Kim et al. (2016), Kotsiantis et al. (2013), Lei et al. (2014), Lonn et al. (2015), Ma et al. (2015), McCormick (2013), Mouri et al. (2015), Munoz-Merino et al. (2015), Ognjanovic et al. (2016), Orduna et al. (2014), Rienties & Toetenel (2016), Rodríguez-Triana et al. (2013), Rodríguez-Triana et al. (2015), Ruiperez-Valiente et al. (2015), Serrano-Laguna et al. (2014), Strang (2016), Tabuenca et al. (2015), Tempelaar et al. (2015), Thompson et al. (2013), Tobarra et al. (2014), Toetenel & Rienties (2016), Vahdat et al. (2016), Van Leeuwen et al. (2015), Yasmin (2013), Yen et al. (2015), You (2016), Zacharis (2015), Zorilla et al. (2015)
						✓	de Barba et al. (2016), Leony et al. (2015), Vu et al. (2015), Xing et al. (2016)
				✓	✓		de Laat & Schreurs (2013), Johri & Teo (2013), Nistor et al. (2014)
			✓	✓			Florian-Gaviria et al. (2013)
			✓	✓	✓		Agudo-Peregrina et al. (2014)

Data Clients

The target population, or *data clients*, are "the beneficiaries of the learning analytics process who are entitled and meant to act upon the outcome (e.g. teachers)" (Drachsler & Greller, 2012, p. 120). Figure 15 shows the distribution of the data clients whom were the target of the learning analytics. Institutions make up the largest group (36 articles), which is commensurate with the high number of articles that present predictive analytics and for the educational setting of higher education being most prominent. Educators (30 articles), in higher education for the most part, are the second most targeted group of clients. Surprisingly only 12 of the articles presented learning analytics targeted for learners. Five articles carried out studies that targeted researchers. Table 21 lists the data clients per article.

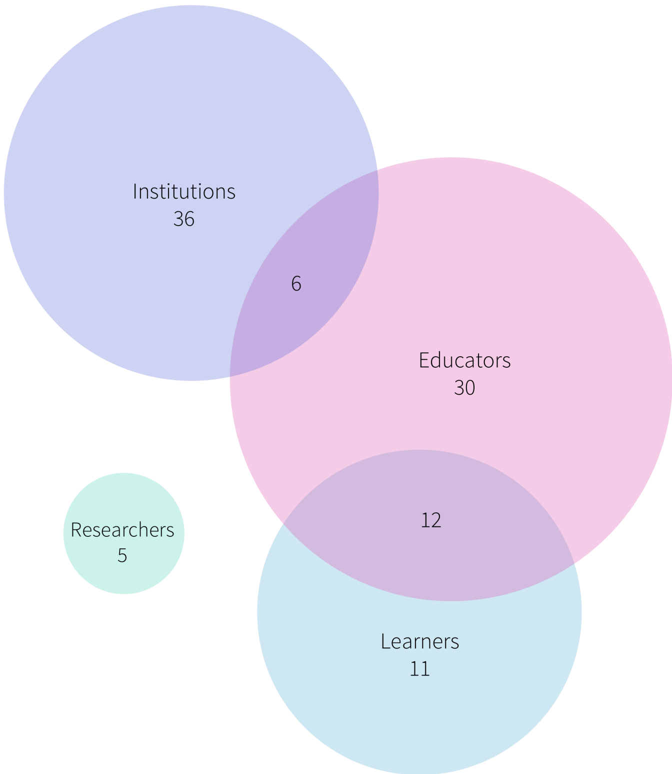


Figure 15 Venn diagram of the data clients

Table 21 Data Clients by article

Citation	Educators	Institutions	Learners	Researchers
Berland et al. (2013), Berland et al. (2014), Berland et al. (2015), Bilkstein et al. (2014, Brooks et al. (2014), Fidalgo-Blanco et al. (2015), Florian-Gaviria et al. (2013), Goggins et al. (2015), Hernandez-Garcia et al. (2015), Hernandez-Nanclares et al. (2016), Jo et al. (2015), Joksimovic et al. (2015b), Lockyer et al. (2013), Lonn et al. (2015), Ma et al. (2015), McCormick (2013), McKenney & Mor (2015), Nistor et al. (2014), Orduna et al. (2014), Persico & Pozzi (2015), Rodriguez-Triana et al. (2013), Rodriguez-Triana et al. (2015), Tempelaar et al. (2015), Van Leeuwen et al. (2014), Van Leeuwen et al. (2015), Vu et al. (2015), Yen et al. (2015), You (2016), Zacharis (2015), Zorilla et al. (2015)	✓			
Bull & Kay (2016), Calvert (2014), Cambruzzi et al. (2015), Casquero et al. (2014), Casquero et al. (2016), Dawson & Siemens (2014), de Barba et al. (2016), de Freitas et al. (2015), de Laat & Prinsen (2014), de Laat & Schreurs (2013), Dix & Leavesley (2015), Drachsler & Kalz (2016), Ferreira & Andrade (2014), Gašević et al. (2013), Ifenthaler & Widanapathirana (2014), Johri & Teo (2013), Kotsiantis et al. (2013), Macfayden et al. (2014), Ognjanovic et al. (2016), Pardo & Siemens (2014), Pardos et al. (2016), Prinsloo & Slade (2014), Prinsloo et al. (2015), Rienties & Toetenel (2016), Robert-Mahoney et al. (2016), Rubel & Jones (2016), Santos et al. (2015), Slade & Prinsloo (2013), Thompson & Cook (2016), Tobarra et al. (2014), Toetenel & Rienties (2016), Volk et al. (2015), Williams (2014), Williamson (2015), Williamson (2016), Yasmin (2013)		✓		
Gibson & de Freitas (2016), Kennedy et al. (2013), Kim et al. (2016), Lei et al. (2014), Martin et al. (2015), Mouri et al. (2015), Tabuenca et al. (2015), Thompson et al. (2013), Xie et al. (2014a), Xie et al. (2014b), Xing et al. (2015)			✓	
Al-Shmoery et al. (2015), Ellis (2013), Lundie (2015), Siemens (2013), Verbert et al. (2014)				✓
Agudo-Peregrina et al. (2014), Akhtar et al. (2015), Joksimovic et al. (2015a), Munoz-Merino et al. (2015), Strang (2016), Xing et al. (2016)	✓	✓		
del Puerto Paule-Ruiz et al. (2015), Fulantelli et al. (2015), Halverson & Owen (2014), Haya et al. (2015), Kickmeier-Rust et al. (2014), Leony et al. (2015), Liu et al. (2016), Melero et al. (2015), Papamitsiou & Economides (2015), Ruiperez-Valiente et al. (2015), Serrano-Laguna et al. (2014), Vahdat et al. (2016)	✓		✓	

Data Subjects

The participants, or *data subjects,* are “the suppliers of data, normally through their browsing and interaction behaviour (e.g. learners) (Drachsler & Greller, 2012, p. 120). In many of the articles it was possible to identify these. In our corpus there is a wide spread number of participants in the studies, see table 22 and figure 16, reported in 67 articles, ranging from less than 50 to over 50,000. Table 22 shows that 32% of the studies had from 100-500 participants, 16% had less than 50 participants, while 7 studies, or 12%, had over 50,000 participants, most likely the MOOC studies.

There were a few articles that made it difficult to place their studies in one or another category. This includes Goggins et al. (2015) who analysed 28 groups of 3-5 members, which means the participants numbered between 84 - 140, meaning it could be place in one of two categories (50 - 100 or 100 - 500), and Toetenel & Rienties (2016) who analysed 157 learning designs where 60,000 students participated, but where the learning designs were analysed and not the student log or demographic data.

Table 22 Aggregated number of data subjects

Study Participants	Freq	%	Size	Freq	%
Learners	56	84%	50 <	9	16%
			50 – 100	9	16%
			100 – 500	18	32%
			500 – 1,000	6	11%
			1,000 – 10,000	5	9%
			10,000 – 50,000	2	4%
			>50,000	7	12%
Educators	4	6%			
Learners + Educators	7	10%			
Total	67				

Instead of specifying specifically the number of participants from which data was collected, some articles wrote about the number of courses with no mention of the number of learners. These include, Ma et al. (2015) that describe the analysis of log data from 900 courses (no mention of the number of learners), and Joksimovic et al. (2015a) that present an analysis of log data from 29 courses (no mention of the number of learners).

Another interesting observation, see table 22, is that in 56 of the 67 studies, 84%, the study participants were learners, and in only 4 studies, 6%, were the participants educators, and 7 studies, 10%, included both leaners and educators.

Table 23 lists participant types by article.

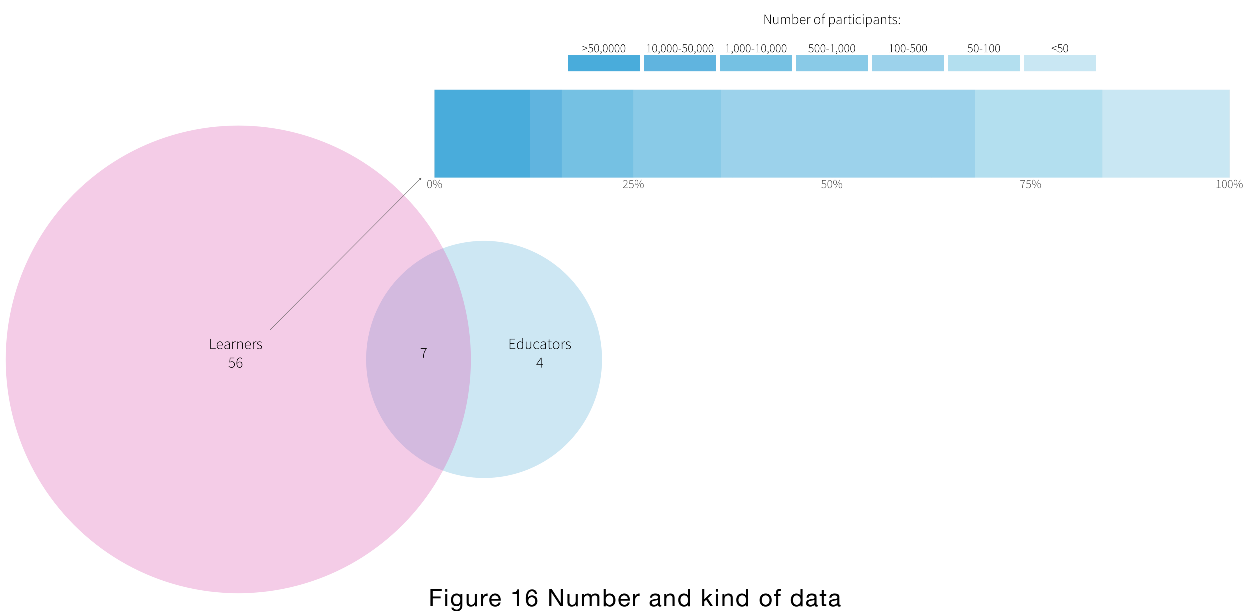


Figure 16 Number and kind of data

Table 23 Data Subjects by article

Size Group	Citation	Total Size	
		Learners	Educators
<50	Thompson (2013)	4	
	Kennedy et al. (2013)	6	
	Liu et al. (2016)	6	
	Xie et al. (2014a)	10	
	Mouri et al. (2015)	17	
	del Puerto Paule-Ruiz et al. (2015)	33	
	Tabuenca et al. (2015)	36	
	Haya et al. (2015)	40	
	Kickmeier-Rust et al. (2014)	40	
50 – 100	Van Leeuwen et al. (2015)	50	
	de Laat & Schreurs (2013)	52	
	Berland et al. (2013)	53	
	Serrano-Laguna et al. (2014)	56	
	Casquero et al. (2014)	61	
	Xie et al. (2014b)	65	
	Tobarra et al. (2014)	68	
	Joksimovic et al. (2015b)	81	
	Leony et al. (2015)	90	
100 – 500	Jo et al. (2015)	100	
	Fidalgo-Blanco et al. (2015)	110	
	Halverson & Owen (2014)	110	
	Casquero et al. (2016)	121	
	Xing et al. (2015)	122	
	Hernandez-Nanclares et al. (2016)	126	
	Nistor et al. (2014)	133	
	Zacharis (2015)	134	
	Lei et al. (2014)	147	
	Kim et al. (2016)	151	
	Lonn et al. (2015)	216	
	Strang (2016)	228	
	Akhtar et al. (2015)	331	
	Kotsiantis et al. (2013)	337	
	Agudo-Peregrina et al. (2014)	356	
	Bilkstein et al. (2014)	370	
	Ruiperez-Valiente et al. (2015)	372	
	Orduna et al. (2014)	433	
500 – 1,000	Gasevic et al. (2013)	505	
	McKenney & Mor (2015)	510	
	You (2016)	530	
	Vahdat et al. (2016)	606	
	de Barba et al. (2016)	862	
	Tempelaar et al. (2015)	922	
1,000 – 10,000	Ognjanovic et al. (2016)	1,061	
	Brooks et al. (2014)	1,379	
	Cambruzzi et al. (2015)	2,491	
	Martin et al. (2015)	3,024	
	Xing et al. (2016)	3,617	
10,000 – 50,000	Yasmin (2013)	12,148	
	Johri & Teo (2013)	21,509	
>50,000	de Freitas et al. (2015)	51,181	
	Gibson & de Freitas (2016)	55,945	
	Vu et al. (2015)	66,286	
	Rientes & Toetenel (2016)	111,256	
	Ifenthaler & Widanapathirana (2014)	158,003	
	Calvert (2014)	60000 – 100000	
	Volk et al. (2015)	>150.000/per month	
	McCormick (2013)		3
	Florian-Gaviria et al. (2013)		20
	Van Leeuwen et al. (2014)		28
	Papamitsiou & Economides (2015)		44
	Rodriguez-Triana et al. (2013)	13	1
	Melero et al. (2015)	81	7
	Berland et al. (2015)	95	3
	Zorilla et al. (2015)	102	2
	Munoz-Merino et al. (2015)	372	8
	Hernandez-Garcia et al. (2015)	656	11
	Yen et al. (2015)	869	3

Pedagogical Approach

Another interesting aspect is the pedagogical approach reported in the articles. Figure 17 shows the frequency of the pedagogical approaches. Collaborative learning studies were by far the most frequent, reported in 24 of the 90 studies (25%). Other examples include, social learning and self-directed/self-regulated learning which were each reported in 10 articles, group learning in 8, networked learning in 5, adaptive or personalised learning in 4, play-based learning in 1. Table 24 lists the articles that use the pedagogical approaches.

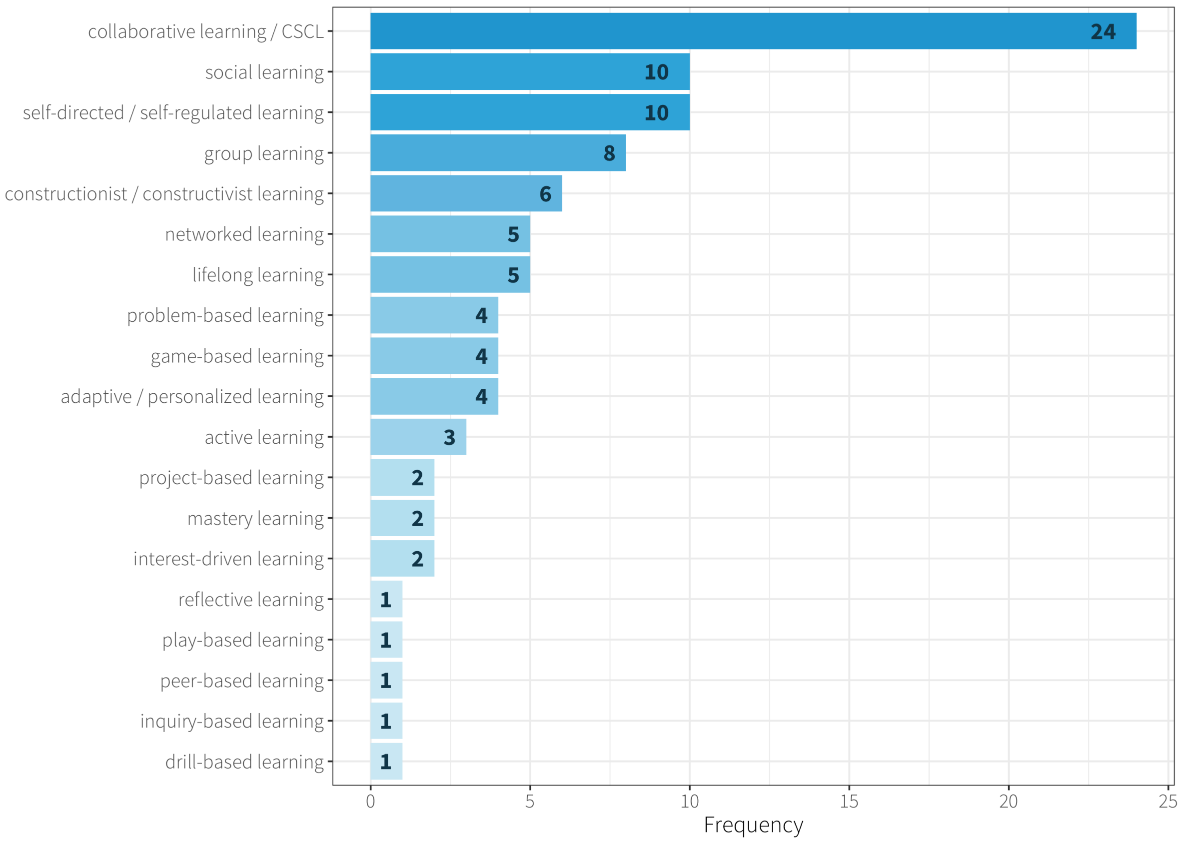


Figure 17 Pedagogical approach by frequency

Table 24 Pedagogical Approach by article

Active Learning	Adaptive / Personalized Learning	Collaborative Learning / CSCL	Constructionist / Constructivist Learning	Drill-based Learning	Game-based Learning	Group Learning	Inquiry-based Learning	Interest-driven Learning	Lifelong Learning	Mastery Learning	Networked Learning	Peer-based Learning	Play-based Learning	Problem-based Learning	Project-based Learning	Reflective Learning	Self-directed/Self-regulated Learning	Social Learning	Citation
✓																			Melero et al. (2015)
	✓																		de Freitas et al. (2015), Leony et al. (2015)
		✓																	Mouri et al. (2015), Rodriguez-Triana et al. (2015), Tobarra et al. (2014), Van Leeuwen et al. (2014), Van Leeuwen et al. (2015), Volk et al. (2015), Yen et al. (2015), Zorilla et al. (2015)
			✓																Berland et al. (2013)
					✓														Gibson & de Freitas (2016), Kickmeier-Rust et al. (2014), Serrano-Laguna et al. (2014)
								✓											Johri & Teo (2013)
									✓										Agudo-Peregrina et al. (2014)
										✓									Lonn et al. (2015)
														✓					Xie et al. (2014a), Xie et al. (2014b)
																✓			Strang (2016)
																	✓		del Puerto Paule-Ruiz et al. (2015), Jo et al. (2015), Lei et al. (2014), Kim et al. (2016), Toetenel & Rienties (2016), Yasmin (2013), You (2016)
																		✓	Nistor et al. (2014), Orduna et al. (2014), Xing et al. (2016)
✓		✓																	Ma et al. (2015)
	✓		✓																Vahdat et al. (2016)
		✓	✓																Berland et al. (2015)
		✓				✓													Akhtar et al. (2015), Fidalgo-Blanco et al. (2015), Goggins et al. (2015), Hernandez-Nanclares et al. (2016), Xing et al. (2015)
		✓							✓										Casquero et al. (2014)
		✓									✓								Thompson et al. (2013)
		✓												✓					Liu et al. (2016)
		✓															✓		Zacharis (2015)
			✓												✓				Berland et al. (2014)
			✓															✓	Rienties & Toetenel (2016)
							✓								✓				Bilkstein et al. (2014)
									✓		✓								de Laat & Schreurs (2013)
									✓								✓		Tabuenca et al. (2015)
										✓							✓		de Barba et al. (2016)
	✓					✓								✓					Tempelaar et al. (2015)
		✓	✓															✓	Kotsiantis et al. (2013)
		✓				✓												✓	Hernandez-Garcia et al. (2015)
		✓							✓									✓	Casquero et al. (2016)
						✓					✓							✓	Gasevic et al. (2013)
✓		✓									✓							✓	Haya et al. (2015)
		✓									✓	✓						✓	Vu et al. (2015)
				✓	✓			✓					✓						Halverson & Owen (2014)

Learning Environment

In some instances it was possible to identify the learning environment in which learning was taking place in the studies. These include: blended, face-to-face, immersive environments (including games and 3D environments), mobile learning, and online learning environments. Table 25 lists the learning environment by article. Figure 18 shows the prevalence of the environments in the studies, with online learning environments being the most popular (in 39 articles), followed by blended learning and immersive environments (10 articles each). A few studies used a mixture of environments such as immersive mobile (2), online and blended (2), online and face-to-face (2), online and mobile (1), online immersive (1), and online, blended, and immersive (1).

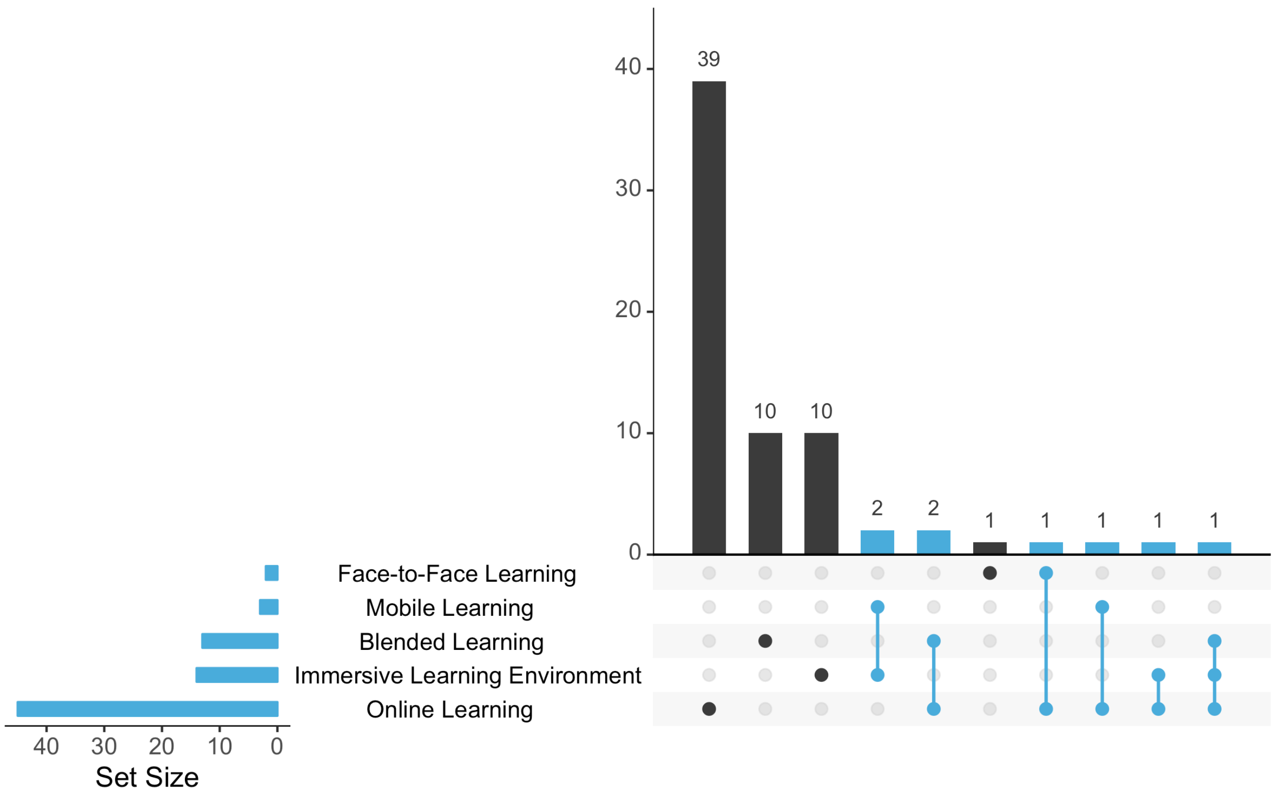


Figure 18 Sets and intersections of the learning environments

Table 25 Learning environments by article

Blended Learning	Face-to-Face Learning	Immersive Learning Environment	Mobile Learning	Online Learning	Citation
✓					Brooks et al. (2014), de Freitas et al. (2015), Fidalgo-Blanco et al. (2015), Hernandez-Nanclares et al. (2016), Kotsiantis et al. (2013), Lonn et al. (2015), Rodríguez-Triana et al. (2013), Rodríguez-Triana et al. (2015), Tempelaar et al. (2015), Zacharis (2015)
	✓				Bilkstein et al. (2014)
		✓			Berland et al. (2013), Berland et al. (2015), Gibson & de Freitas (2016), Goggins et al. (2015), Halverson & Owen (2014), Kennedy et al. (2013), Kickmeier-Rust et al. (2014), Liu et al. (2016), Serrano-Laguna et al. (2014), Xie et al. (2014a), Xie et al. (2014b)
				✓	Akhtar et al. (2015), Calvert (2014), Cambruzzi et al. (2015), Casquero et al. (2014), Casquero et al. (2016), de Barba et al. (2016), de Laat & Schreurs (2013), del Puerto Paule-Ruiz et al. (2015), Gašević et al. (2013), Haya et al. (2015), Hernandez-Garcia et al. (2015), Ifenthaler & Widanapathirana (2014), Jo et al. (2015), Johri & Teo (2013), Joksimovic et al. (2015a), Joksimovic et al. (2015b), Kim et al. (2016), Lei et al. (2014), Leony et al. (2015), Ma et al. (2015), McCormick (2013), Munoz-Merino et al. (2015), Nistor et al. (2014), Ognjanovic et al. (2016), Papamitsiou & Economides (2015), Rienties & Toetenel (2016), Ruiperez-Valiente et al. (2015), Strang (2016), Tobarra et al. (2014), Van Leeuwen et al. (2014), Van Leeuwen et al. (2015), Volk et al. (2015), Vu et al. (2015), Xing et al. (2015), Xing et al. (2016), Yen et al. (2015), You (2016), Zorilla et al. (2015)
✓				✓	Agudo-Peregrina et al. (2014), Florian-Gaviria et al. (2013)
	✓			✓	Thompson et al. (2013)
		✓	✓		Melero et al. (2015), Mouri et al. (2015)
		✓		✓	Orduna et al. (2014)
			✓	✓	Tabuenca et al. (2015)
✓		✓		✓	Martin et al. (2015)

4. Summary

In this state of the field report, we have carried out a search for learning analytics articles from a *pedagogical perspective*, meaning that we have not searched, for example, for algorithms or methods used in learning analytics. We developed a search string that looked for "learning analytics" combined with keywords that address 1) problems in education (e.g., retention, decision support), 2) a particular level of education (K-12, higher education), 3) outcomes (e.g., data literacy, performance), 4) stakeholders (e.g., teachers, learners, policy makers), and 5) implementation settings (e.g., personalised learning, infrastructure).

From the overview of the corpus is evident that the number of articles are growing each year, with an increase from 26 articles in 2014 to 40 articles in 2015, and 20 articles in the first 2 months alone of 2016. In addition, the titles of the 44 journals show that the articles are distributed among computer science, education, and psychology journals.

In an analysis of the corpus of articles we have seen that learning analytics research fall under a wide range of themes. We identified seven themes — *algorithms & models, data, predictive analysis, text analysis, tool development and visualisations* — that deal with the development of methods, tools, or presentations for learning analytics. Two themes — *LA for Institutions,* and *LA for Educators* — were focused on providing learning analytics for a target group either at the macro level for institutions, or at the micro and meso level for educators. One *overview* article that gives a historical perspective on learning analytics was included; the author George Siemens is attributed with the "invention" of the MOOC. One theme, *LA+*, shows that learning analytics is an interdisciplinary effort and learning analytics researchers and researchers in areas such as assessment need to join forces. Finally, it is heartening to see that there are researchers concerned with the *ethics, philosophy, policy theme* as the implementation of learning analytics is challenged by privacy regulations and has an ethical and philosophical dimension.

There is also a wide range of topics addressed in the various themes. From the thematic analysis we see that the *algorithms & models* are focused on understanding learning processes or predicting student performance, mostly using machine learning and educational data mining methods. There is a variety of *data* sources being use, with the most popular being data logs from a variety of learning environments including LMSs, MOOCs, games, immersive environments. Some studies also used interaction data and the content of discussion forum messages.

The articles concerned with *ethics, philosophy, & policy* address issues such as the role of algorithms, analytics as a moral practice, power relations, and various aspects of privacy, autonomy, and datafication. The theoretical papers found in the *LA+* theme aim at marrying LA with an established discipline such as mobile learning, MOOC research, or assessment.

The *LA for educators* theme looks at LA for learning or curriculum design, detecting low performing groups and to support teacher inquiry. *LA for institutions*, on the other hand, is focused on academic challenges such as dropouts, retention, as well as improving assessment and supporting decisions in context.

The *network analysis* theme uses various aspects of social network analysis to analyse or predict aspects of learner behaviour in online learning communities, personal learning environments or informal networks, to understand interaction patterns and structures. The *predictive analysis* theme addressed issues such as predicting dropout rates, dropouts, academic performance, and student success. The *Text analysis* theme covered research on using learning analytics to analyse essays or discussion forum messages. This research is attempting to get to learning-centric analytics, which is focussed on understanding learning by looking at artefacts generated by learners.

The tool development theme introduced learning analytics tools to assist with formative feedback, self-regulated learning, visualisation of learner models and course models. Finally, the visualisations theme covered research on using learning analytics to help the visualisation of learning and learning processes, and whether or not teachers and students could understand the visualisations presented to them, including in dashboards.

From our analysis we see that there are a number of different data sources and methods being used, but it is clear that the most common of the 9 data source is *log data*, followed by a combination of *log data and performance data* (e.g., course marks), and the most common of the 24 analysis method is *descriptive statistics*, followed by *correlation analysis* and *regression analysis*.

The educational setting most seen in the studies is *higher education*, with 48 of the 63 identifiable settings being universities. There are only 5 studies addressing *high school* settings, and none addressing primary or middle school.

The majority of the contributions from the articles were directed to the *application of models in various educational settings*, followed by *theoretical contributions*, and the *application of developed tools in various educational settings*.

The target audience, or *data clients, was most often institutions*, followed by *educators*, and the majority of the *participants, or data subjects, were learners*. The number of learners in the study ranged widely from less that 50 to more than 50,000, with the average study having between 100 and 500 participants.

Computer supported collaborative learning (CSCL) / collaborative learning was the *pedagogical approach* most often found in the studies. This is not surprising as the CSCL community is mature, and has been using social network analysis and interaction analysis to understand interaction between learners, between learners and teachers and between learners and learning resources for over two decades. This also explains the prevalence of social learning approaches. Similarly, the prevalence of *self-regulated learning* (SRL) can be explained as SRL is also being a mature field, where in recent years they have been looking at how new types of data, such as biosensors or eye-tracking, can help explain SRL.

Finally, it is not surprising that online learning environments were over three times more frequently studied than blended learning or immersive environments. The availability of log data from online learning situations makes it easier to apply learning analytics, and this corresponds with log data being the most prevalent data source.

5. Conclusions

Learning analytics is an emerging, fast growing field of research. From the first articles that appeared at the 2010 conference to the growing corpus of journal articles, we see an exciting field of research that has the potential to impact education on many levels.

In conclusion, there are a number of general observations that can be gleaned from this state of the field study:

- an analysis of the corpus shows evidence that learning analytics is a broad field with articles published in education, computer science, and psychology journals, and there are wide range of research topics addressed in the research, both technical and educational
- the definition of "learning analytics" is still under discussion; some make a division between academic analytics and learning analytics, but as this corpus illustrates, many use learning analytics to cover academic analytics
- the research is data rich, but theory poor; studies often lack theoretical, historical or pedagogical perspectives
- there is a predominance of studies in higher education, informal learning, and distance education settings, with a few studies in high school, and none in primary or middle school
- predictive analysis is a very popular research area addressing HE problems such as dropouts, retention, and curriculum issues
- predictive models are situation dependent and there is little evidence that they are transferable to other contexts
- privacy issues are rarely addressed in the case studies, but there are papers that do focus on ethics and privacy
- there are very few implementation studies and impact studies (this supports Ferguson et al.'s (2016)[14] finding), indicating that as of yet, learning analytics is an immature field

There are also a number of gaps in the research on learning analytics:

- the application of learning analytics in K-12 education (at macro, meso, micro levels)
- research on everyday analytics in classrooms (i.e., how do we collect data in classrooms)
- research on assessment/feedback
- research on learning-centric analytics, as opposed to learner-centric (*cf* Stein, 2012)[15] meaning we need also a focus learning and learning outcomes and not just learner behaviour
- implementation and impact of learning analytics
- data literacy, although there are a few studies addressing whether or not stakeholders can understand the visualisations they are presented.

Carrying out learning analytics research is not easy. You need an interdisciplinary team with a common vision. There are challenges such as access to data, storage of data, finding the most useful

[14] Ferguson, R., Brasher, A., Clow, D., Cooper, A., Hillaire, G., Mittelmeier, J., Rienties, B., Ullmann, T., Vuorikari, R. (2016). Research Evidence on the Use of Learning Analytics - Implications for Education Policy. R. Vuorikari, J. Castaño Muñoz (Eds.). *Joint Research Centre Science for Policy Report;* EUR 28294 EN; doi:10.2791/955210.

[15] Stein, Z. (2012). Learning Analytics and the Learning Sciences, *ELI Webinar,* August 13, 2012, http://www.educause.edu/events/eli-webinar- learning-analytics-and-learning-sciences.

analytics methods and selecting the right data, accessing learning and learning processes and not just behaviour, and being aware of the ethics of algorithm bias, and privacy aspects that need to be addressed thoughtfully.

Despite this, learning analytics is an exciting and important field of research, which will continue to grow in the coming years. As this report shows, researchers in the field are concerned with not only the technical aspects of learning analytics, but also with the pedagogical and psychological, and ethical and philosophical perspectives. As the field is still immature and there are little implementation studies and impact outcomes, it is difficult to give advice to policy makers on what works, but there is promise in this approach and policy makers should support research in this area.

References

Algorithms & Models

Berland et al. (2013)

Berland, M., Martin, T., Benton, T., Petrick Smith, C., & Davis, D. (2013). Using Learning Analytics to Understand the Learning Pathways of Novice Programmers. Journal of the Learning Sciences, 22(4), 564-599.

Brooks et al. (2014)

Brooks, C., Erickson, G., Greer, J., & Gutwin, C. (2014). Modelling and Quantifying the Behaviours of Students in Lecture Capture Environments. Computers & Education, 75, 282-292.

del Puerto Paule-Ruiz et al. (2015)

del Puerto Paule-Ruiz, M., Riestra-Gonzalez, M., Sánchez-Santillan, M., & Pérez-Pérez, J. R. (2015). The Procrastination Related Indicators in e-Learning Platforms. Journal of Universal Computer Science, 21(1), 7-22.

Goggins et al. (2015)

Goggins, S. P., Xing, W., Chen, X., Chen, B., & Wadholm, B. (2015). Learning Analytics at "Small" Scale: Exploring a Complexity-Grounded Model for Assessment Automation. Journal of Universal Computer Science, 21(1), 66-92.

Joksimovic et al. (2015a)

Joksimović, S., Gašević, D., Loughin, T. M., Kovanović, V., & Hatala, M. (2015). Learning at Distance: Effects of Interaction Traces on Academic Achievement. Computers & Education, 87, 204-217.

Leony et al. (2015)

Leony, D., Merino, P. J. M., Valiente, J. A. R., Pardo, A., & Kloos, C. D. (2015). Detection and Evaluation of Emotions in Massive Open Online Courses. Journal of Universal Computer Science, 21(5), 638-655.

Ma et al. (2015)

Ma, J., Han, X., Yang, J., & Cheng, J. (2015). Examining the Necessary Condition for Engagement in an Online Learning environment Based on Learning Analytics Approach: The Role of the Instructor. The Internet and Higher Education, 24, 26-34.

Martin et al. (2015)

Martin, T., Petrick Smith, C., Forsgren, N., Aghababyan, A., Janisiewicz, P., & Baker, S. (2015). Learning Fractions by splitting: Using Learning Analytics to Illuminate the Development of Mathematical Understanding. Journal of the Learning Sciences, 24(4), 593-637.

Munoz-Merino et al. (2015)

Muñoz-Merino, P. J., Ruipérez-Valiente, J. A., Alario-Hoyos, C., Pérez-Sanagustín, M., & Kloos, C. D. (2015). Precise Effectiveness Strategy for Analyzing the Effectiveness of Students with Educational Resources and Activities in MOOCs. Computers in Human Behavior, 47, 108-118.

Nistor et al. (2014)

Nistor, N., Baltes, B., Dascălu, M., Mihăilă, D., Smeaton, G., & Trăuşan-Matu, Ş. (2014). Participation in Virtual Academic Communities of Practice under the influence of Technology Acceptance and Community Factors. A Learning Analytics Application. Computers in Human Behavior, 34, 339-344.

Vahdat et al. (2016)

Vahdat, M., Oneto, L., Anguita, D., Funk, M., & Rauterberg, M. (2016). Can Machine Learning Explain Human Learning?. Neurocomputing, 192, 14-28.

Xing et al. (2015)

Xing, W., Guo, R., Petakovic, E., & Goggins, S. (2015). Participation-based Student Final Performance Prediction Model Through Interpretable Genetic Programming: Integrating Learning Analytics, Educational Data Mining and Theory. Computers in Human Behavior, 47, 168-181.

Data

Gibson & de Freitas (2016)

Gibson, D., & de Freitas, S. (2016). Exploratory Analysis in Learning Analytics. Technology, Knowledge and Learning, 21(1), 5-19.

Halverson & Owen (2014)

Halverson, R., & Owen, V. E. (2014). Game-based Assessment: An Integrated Model for Capturing Evidence of Learning in Play. International Journal of Learning Technology, 9(2), 111-138.

Kennedy et al. (2013)

Kennedy, G., Ioannou, I., Zhou, Y., Bailey, J., & O'Leary, S. (2013). Mining Interactions in Immersive Learning Environments for Real-Time Student Feedback. Australasian Journal of Educational Technology, 29(2).

Liu et al. (2016)

Liu, M., Lee, J., Kang, J., & Liu, S. (2016). What We can Learn from the Data: A Multiple-case Study Examining Behavior Patterns by Students with Different Characteristics in Using a Serious Game. Technology, Knowledge and Learning, 21(1), 33-57.

Santos et al. (2015)

Santos, J. L., Verbert, K., Klerkx, J., Duval, E., Charleer, S., & Ternier, S. (2015). Tracking Data in Open Learning Environments. Journal of Universal Computer Science, 21(7), 976-996.

Serrano-Laguna et al. (2014)

Serrano-Laguna, Á., Torrente, J., Moreno-Ger, P., & Fernández-Manjón, B. (2014). Application of Learning Analytics in Educational Videogames. Entertainment Computing, 5(4), 313-322.

Thompson et al. (2013)

Thompson, K., Ashe, D., Carvalho, L., Goodyear, P., Kelly, N., & Parisio, M. (2013). Processing and Visualizing Data in Complex Learning Environments. American Behavioral Scientist, 57(10), 1401-1420.

Xie et al. (2014a)

Xie, C., Zhang, Z., Nourian, S., Pallant, A., & Hazzard, E. (2014). A Time Series Analysis Method for Assessing Engineering Design Processes Using a CAD Tool. International Journal of Engineering Education, 30(1), 218-230.

Xie et al. (2014b)

Xie, C., Zhang, Z., Nourian, S., Pallant, A. M. Y., & Bailey, S. (2014). On the Instructional Sensitivity of CAD Logs. International Journal of Engineering Education, 30(4), 760-778.

Yen et al. (2015)

Yen, C. H., Chen, I. C., Lai, S. C., & Chuang, Y. R. (2015). An Analytics-Based Approach to Managing Cognitive Load by Using Log Data of Learning Management Systems and Footprints of Social Media. Journal of Educational Technology & Society, 18(4), 141-158.

Ethics, Philosophy & Policy

Lundie (2015)

Lundie, D. (2017). The Givenness of the Human Learning Experience and its Incompatibility with Information Analytics. Educational Philosophy and Theory, 49(4), 391-404.[16]

Pardo & Siemens (2014)

Pardo, A., & Siemens, G. (2014). Ethical and Privacy Principles for Learning Analytics. British Journal of Educational Technology, 45(3), 438-450.

Prinsloo & Slade (2014)

Prinsloo, P., & Slade, S. (2014). Educational Triage in Open Distance Learning: Walking a Moral Tightrope. The International Review Of Research In Open And Distance Learning, 15(4), 306-331.

Roberts-Mahoney et al. (2016)

[16] Published online in 2015

Roberts-Mahoney, H., Means, A. J., & Garrison, M. J. (2016). Netflixing Human Capital Development: Personalized Learning Technology and the Corporatization of K-12 Education. Journal of Education Policy, 31(4), 405-420.

Rubel & Jones (2016)

Rubel, A., & Jones, K. M. (2016). Student Privacy in Learning Analytics: An Information Ethics Perspective. The Information Society, 32(2), 143-159.

Slade & Prinsloo (2013)

Slade, S., & Prinsloo, P. (2013). Learning Analytics: Ethical Issues and Dilemmas. American Behavioral Scientist, 57(10), 1510-1529.

Thompson & Cook (2016)

Thompson, G., & Cook, I. (2016). The Logic of Data-Sense: Thinking through Learning Personalisation. Discourse: Studies in the Cultural Politics of Education, 1-15.

Williamson (2015)

Williamson, B. (2015). Governing Software: Networks, Databases and Algorithmic Power in the Digital Governance of Public Education. Learning, Media and Technology, 40(1), 83-105.

Williamson (2016)

Williamson, B. (2016). Digital Education Governance: Data Visualization, Predictive Analytics, and 'Real-Time' Policy Instruments. Journal of Education Policy, 31(2), 123-141.

LA+

Berland et al. (2014)

Berland, M., Baker, R. S., & Blikstein, P. (2014). Educational Data Mining and Learning Analytics: Applications to Constructionist Research. Technology, Knowledge and Learning, 19(1-2), 205-220.

Drachsler & Kalz (2016)

Drachsler, H., & Kalz, M. (2016). The MOOC and Learning Analytics Innovation Cycle (MOLAC): A Reflective Summary of Ongoing Research and its Challenges. Journal of Computer Assisted Learning, 32(3), 281-290.

Ellis (2013)

Ellis, C. (2013). Broadening the Scope and Increasing the Usefulness of Learning Analytics: The Case for Assessment Analytics. British Journal of Educational Technology, 44(4), 662-664.

Fulantelli et al. (2015)

Fulantelli, G., Taibi, D., & Arrigo, M. (2015). A Framework to Support Educational Decision Making in Mobile Learning. Computers in Human Behavior, 47, 50-59.

Williams (2014)

Williams, P. (2014). Squaring the Circle: A New Alternative to Alternative-Assessment. Teaching in Higher Education, 19(5), 565-577.

LA for Educators

Berland et al. (2015)

Berland, M., Davis, D., & Smith, C. P. (2015). AMOEBA: Designing for Collaboration in Computer Science Classrooms through Live Learning Analytics. International Journal of Computer-Supported Collaborative Learning, 10(4), 425-447.

Florian-Gaviria et al. (2013)

Florian-Gaviria, B., Glahn, C., & Gesa, R. F. (2013). A Software Suite for Efficient Use of the European Qualifications Framework in Online and Blended Courses. IEEE Transactions on Learning Technologies, 6(3), 283-296.

Haya et al. (2015)

Haya, P. A., Daems, O., Malzahn, N., Castellanos, J., & Hoppe, H. U. (2015). Analysing Content and Patterns of Interaction for Improving the Learning Design of Networked Learning Environments. British Journal of Educational Technology, 46(2), 300-316.

Lockyer et al. (2013)
Lockyer, L., Heathcote, E., & Dawson, S. (2013). Informing Pedagogical Action: Aligning Learning Analytics with Learning Design. American Behavioral Scientist, 57(10), 1439-1459.

McKenney & Mor (2015)
McKenney, S., & Mor, Y. (2015). Supporting Teachers in Data-informed Educational Design. British Journal of Educational Technology, 46(2), 265-279.

Persico & Pozzi (2015)
Persico, D., & Pozzi, F. (2015). Informing Learning Design with Learning Analytics to Improve Teacher Inquiry. British Journal of Educational Technology, 46(2), 230-248.

Rodriguez-Triana et al. (2013)
Rodríguez-Triana, M. J., Martínez-Monés, A., Asensio-Pérez, J. I., & Dimitriadis, Y. (2013). Towards a Script-aware Monitoring Process of Computer-supported Collaborative Learning Scenarios. International Journal of Technology Enhanced Learning, 5(2), 151-167.

Rodriguez-Triana et al. (2015)
Rodríguez-Triana, M. J., Martínez-Monés, A., Asensio-Pérez, J. I., & Dimitriadis, Y. (2015). Scripting and Monitoring Meet Each Other: Aligning Learning Analytics and Learning Design to Support Teachers in Orchestrating CSCL Situations. British Journal of Educational Technology, 46(2), 330-343.

Van Leeuwen et al. (2014)
Van Leeuwen, A., Janssen, J., Erkens, G., & Brekelmans, M. (2014). Supporting Teachers in Guiding Collaborating Students: Effects of Learning Analytics in CSCL. Computers & Education, 79, 28-39.

Van Leeuwen et al. (2015)
Van Leeuwen, A., Janssen, J., Erkens, G., & Brekelmans, M. (2015). Teacher Regulation of Cognitive Activities During Student Collaboration: Effects of Learning Analytics. Computers & Education, 90, 80-94.

LA for Institutions

Dawson & Siemens (2014)
Dawson, S., & Siemens, G. (2014). Analytics to Literacies: The Development of a Learning Analytics Framework for Multiliteracies Assessment. The International Review of Research in Open and Distributed Learning, 15(4), 284-305.

de Freitas et al. (2015)
Freitas, S., Gibson, D., Du Plessis, C., Halloran, P., Williams, E., Ambrose, M., ... & Arnab, S. (2015). Foundations of Dynamic Learning Analytics: Using University Student Data to Increase Retention. British Journal of Educational Technology, 46(6), 1175-1188.

de Laat & Prinsen (2014)
De Laat, M., & Prinsen, F. R. (2014). Social Learning Analytics: Navigating the Changing Settings of Higher Education. Research & Practice in Assessment, 9, 51-60.

Dix & Leavesley (2015)
Dix, A. J., & Leavesley, J. (2015). Learning Analytics for the Academic: An Action Perspective. Journal of Universal Computer Science, 21(1), 48-65.

Ferreira & Andrade (2014)
Ferreira, S. A., & Andrade, A. (2014). Academic Analytics: Mapping the Genome of the University. IEEE Revista Iberoamericana de Tecnologias del Aprendizaje, 9(3), 98-105.

Macfayden et al. (2014)
Macfadyen, L. P., Dawson, S., Pardo, A., & Gasevic, D. (2014). Embracing Big Data in Complex Educational Systems: The Learning Analytics Imperative and the Policy Challenge. Research & Practice in Assessment, 9.

Prinsloo et al. (2015)

Prinsloo, P., Archer, E., Barnes, G., Chetty, Y., & Van Zyl, D. (2015). Big(ger) Data as Better Data in Open Distance Learning. The International Review of Research in Open and Distributed Learning, 16(1), 284-306.

Rientes & Toetenel (2016)

Rienties, B., & Toetenel, L. (2016). The Impact of Learning Design on Student Behaviour, Satisfaction and Performance: A Cross-Institutional Comparison Across 151 Modules. Computers in Human Behavior, 60, 333-341.

Toetenel & Rienties (2016)

Toetenel, L., & Rienties, B. (2016). Analysing 157 Learning Designs Using Learning Analytic Approaches as a Means to Evaluate the Impact of Pedagogical Decision Making. British Journal of Educational Technology, 47(5), 981-992.

Volk et al. (2015)

Volk, H., Kellner, K., & Wohlhart, D. (2015). Learning Analytics for English Language Teaching. Journal of Universal Computer Science, 21(1), 156-174.

Yasmin (2013)

Yasmin, D. (2013). Application of the Classification Tree Model in Predicting Learner Dropout Behaviour in Open and Distance Learning. Distance Education, 34(2), 218-231.

Network Analysis

Casquero et al. (2014)

Casquero, O., Ovelar, R., Romo, J., & Benito, M. (2014). Personal Learning Environments, Higher Education and Learning Analytics: A Study of the Effects of Service Multiplexity on Undergraduate Students' Personal Networks/Entornos de aprendizaje personales, educación superior y analítica del aprendizaje: un estudio sobre los efectos de la multiplicidad de servicios en las redes personales de estudiantes universitarios. Cultura y Educación, 26(4), 696-738.

Casquero et al. (2016)

Casquero, O., Ovelar, R., Romo, J., Benito, M., & Alberdi, M. (2016). Students' Personal Networks in Virtual and Personal Learning Environments: A Case Study in Higher Education Using Learning Analytics Approach. Interactive Learning Environments, 24(1), 49-67.

de Laat & Schreurs (2013)

de Laat, M., & Schreurs, B. (2013). Visualizing Informal Professional Development Networks: Building a Case for Learning Analytics in the Workplace. American Behavioral Scientist, 57(10), 1421-1438.

Hernandez-Garcia et al. (2015)

Hernández-García, Á., González-González, I., Jiménez-Zarco, A. I., & Chaparro-Peláez, J. (2015). Applying Social Learning Analytics to Message Boards in Online Distance Learning: A Case Study. Computers in Human Behavior, 47, 68-80.

Hernandez-Nanclares et al. (2016)

Hernández-Nanclares, N., García-Muñiz, A. S., & Rienties, B. (2017). Making the Most of "External" Group Members in Blended and Online Environments. Interactive Learning Environments, 25(4), 1-15.[17]

Johri & Teo (2013)

Johri, A., & Teo, H. J. (2013). Using Data Analytics to Examine Expert/Novice Behavior in Informal Online Communities. Bulletin of the IEEE Technical Committee on Learning Technology, 15(2), 2.

Mouri et al. (2013)

Mouri, K., Ogata, H., & Uosaki, N. (2015). Analysis of Ubiquitous Learning Logs Using Social Network Analysis. International Journal of Mobile Learning and Organisation, 9(2), 101-123.

Orduna et al. (2014)

[17] Published online in 2016

Orduña, P., Almeida, A., Ros, S., López-de-Ipiña, D., & Zubía, J. G. (2014). Leveraging Non-explicit Social Communities for Learning Analytics in Mobile Remote Laboratories. Journal of Universal Computer Science, 20(15), 2043-2053.

Overview

Siemens (2013)

Siemens, G. (2013). Learning Analytics: The Emergence of a Discipline. American Behavioral Scientist, 57(10), 1380-1400.

Predictive Analysis

Agudo-Peregrina et al. (2014)

Agudo-Peregrina, Á. F., Iglesias-Pradas, S., Conde-González, M. Á., & Hernández-García, Á. (2014). Can we Predict Success from Log Data in VLEs? Classification of Interactions for Learning Analytics and their Relation with Performance in VLE-Supported F2F and Online Learning. Computers in Human Behavior, 31, 542-550.

Akhtar et al. (2015)

Akhtar, S., Warburton, S., & Xu, W. (2017). The use of an Online Learning and Teaching System for Monitoring Computer Aided Design Student Participation and Predicting Student Success. International Journal of Technology and Design Education, 27(2), 251-270.[18]

Blikstein et al. (2014)

Blikstein, P., Worsley, M., Piech, C., Sahami, M., Cooper, S., & Koller, D. (2014). Programming Pluralism: Using Learning Analytics to Detect Patterns in the Learning of Computer Programming. Journal of the Learning Sciences, 23(4), 561-599.

Calvert (2014)

Calvert, C. E. (2014). Developing a Model and Applications for Probabilities of Student Success: A Case Study of Predictive Analytics. Open Learning: The Journal of Open, Distance and e-Learning, 29(2), 160-173.

Cambruzzi et al. (2015)

Cambruzzi, W. L., Rigo, S. J., & Barbosa, J. L. (2015). Dropout Prediction and Reduction in Distance Education Courses with the Learning Analytics Multitrail Approach. Journal of Universal Computer Science, 21(1), 23-47.

de Barba et al. (2016)

De Barba, P. D., Kennedy, G. E., & Ainley, M. D. (2016). The Role of Students' Motivation and Participation in Predicting Performance in a MOOC. Journal of Computer Assisted Learning, 32(3), 218-231.

Fidalgo-Blanco et al. (2015)

Fidalgo-Blanco, Á., Sein-Echaluce, M. L., García-Peñalvo, F. J., & Conde, M. Á. (2015). Using Learning Analytics to Improve Teamwork Assessment. Computers in Human Behavior, 47, 149-156.

Gasevic et al. (2013)

Gašević, D., Zouaq, A., & Janzen, R. (2013). “Choose your Classmates, your GPA is at Stake!” The Association of Cross-Class Social Ties and Academic Performance. American Behavioral Scientist, 57(10), 1460-1479.

Ifenthaler & Widanapathirana (2014)

Ifenthaler, D., & Widanapathirana, C. (2014). Development and Validation of a Learning Analytics Framework: Two Case Studies Using Support Vector Machines. Technology, Knowledge and Learning, 19(1-2), 221-240.

Jo et al. (2015)

Jo, I.H., Kim, D., & Yoon, M. (2015). Constructing Proxy Variables to Measure Adult Learners' Time Management Strategies in LMS. Journal of Educational Technology & Society, 18(3), 214-225.

[18] First online in 2015

Joksimovic et al. (2015b)

Joksimović, S., Gašević, D., Kovanović, V., Riecke, B. E., & Hatala, M. (2015). Social Presence in Online Discussions as a Process Predictor of Academic Performance. Journal of Computer Assisted Learning, 31(6), 638-654.

Kotsiantis et al. (2013)

Kotsiantis, S., Tselios, N., Filippidi, A., & Komis, V. (2013). Using Learning Analytics to Identify Successful Learners in a Blended Learning Course. International Journal of Technology Enhanced Learning, 5(2), 133-150.

Lonn et al. (2015)

Lonn, S., Aguilar, S. J., & Teasley, S. D. (2015). Investigating Student Motivation in the Context of a Learning Analytics Intervention During a Summer Bridge Program. Computers in Human Behavior, 47, 90-97.

Ognjanovic et al. (2016)

Ognjanovic, I., Gasevic, D., & Dawson, S. (2016). Using Institutional Data to Predict Student Course Selections in Higher Education. The Internet and Higher Education, 29, 49-62.

Strang (2016)

Strang, K. D. (2016). Beyond engagement analytics: which online mixed-data factors predict student learning outcomes?. Education and information technologies, 22(3), 917-937.

Tempelaar et al. (2015)

Tempelaar, D. T., Rienties, B., & Giesbers, B. (2015). In Search for the Most Informative Data for Feedback Generation: Learning Analytics in a Data-rich Context. Computers in Human Behavior, 47, 157-167.

Vu et al. (2015)

Vu, D., Pattison, P., & Robins, G. (2015). Relational Event Models for Social Learning in MOOCs. Social Networks, 43, 121-135.

Xing et al. (2016)

Xing, W., Chen, X., Stein, J., & Marcinkowski, M. (2016). Temporal Predication of Dropouts in MOOCs: Reaching the Low Hanging Fruit through Stacking Generalization. Computers in Human Behavior, 58, 119-129.

You (2016)

You, J. W. (2016). Identifying Significant Indicators Using LMS Data to Predict Course Achievement in Online Learning. The Internet and Higher Education, 29, 23-30.

Zacharis (2015)

Zacharis, N. Z. (2015). A Multivariate Approach to Predicting Student Outcomes in Web-enabled Blended Learning Courses. The Internet and Higher Education, 27, 44-53.

Text Analysis

Al-Shmoery et al. (2015)

Al-Shmoery, Yahya, Messoussi, Rochdi, & Touahni, Raja (2015). Analytical Tools for Visualisation of Interactions in Online E-Learning Activities on LMS and Semantic Similarity Measures on Text. Journal of Theoretical & Applied Information Technology, 73(1), 102-118.

Lei et al. (2014)

Lei, C. U., Man, K. L., & Ting, T. O. (2014). Using Learning Analytics to Analyze Writing Skills of Students: A Case Study in a Technological Common Core Curriculum Course. IAENG International Journal of Computer Science, 41(3).

Tobarra et al. (2014)

Tobarra, L., Robles-Gómez, A., Ros, S., Hernández, R., & Caminero, A. C. (2014). Analyzing the Students' Behavior and Relevant Topics in Virtual Learning Communities. Computers in Human Behavior, 31, 659-669.

Tool Development

Bull & Kay (2016)

Bull, S., & Kay, J. (2016). SMILI: A Framework for Interfaces to Learning Data in open Learner Models, Learning Analytics and Related Fields. International Journal of Artificial Intelligence in Education, 26(1), 293-331.

Kickmeier-Rust et al. (2014)

Kickmeier-Rust, M. D., Hillemann, E. C., & Albert, D. (2014). Gamification and Smart Feedback: Experiences with a Primary School Level Math App. International Journal of Game-Based Learning, 4(3), 35-46.

Pardos et al. (2016)

Pardos, Z. A., Whyte, A., & Kao, K. (2016). moocRP: Enabling Open Learning Analytics with an Open Source Platform for Data Distribution, Analysis, and Visualization. Technology, Knowledge and Learning, 21(1), 75-98.

Tabuenca et al. (2015)

Tabuenca, B., Kalz, M., Drachsler, H., & Specht, M. (2015). Time will Tell: The Role of Mobile Learning Analytics in Self-Regulated Learning. Computers & Education, 89, 53-74.

Visualisations

Kim et al. (2016)

Kim, J., Jo, I. H., & Park, Y. (2016). Effects of Learning Analytics Dashboard: Analyzing the Relations Among Dashboard Utilization, Satisfaction, and Learning Achievement. Asia Pacific Education Review, 17(1), 13-24.

McCormick (2013)

McCormick, J. (2013). Visualizing Interaction: Pilot Investigation of a Discourse Analytics Tool for Online Discussion. Bulletin of the IEEE Technical Committee on Learning Technology, 15(2), 10-13.

Melero et al. (2016)

Melero, J., Hernández-Leo, D., Sun, J., Santos, P., & Blat, J. (2015). How was the Activity? A Visualization Support for a Case of Location-based Learning Design. British Journal of Educational Technology, 46(2), 317-329.

Papamitsiou & Economides (2015)

Papamitsiou, Z., & Economides, A. A. (2015). Temporal Learning Analytics Visualizations for Increasing Awareness During Assessment. International Journal of Educational Technology in Higher Education, 12(3), 129-147.

Ruiperez-Valiente et al. (2015)

Ruipérez-Valiente, J. A., Muñoz-Merino, P. J., Leony, D., & Kloos, C. D. (2015). ALAS-KA: A Learning Analytics Extension for Better Understanding the Learning Process in the Khan Academy Platform. Computers in Human Behavior, 47, 139-148.

Verbert et al. (2014)

Verbert, K., Govaerts, S., Duval, E., Santos, J. L., Van Assche, F., Parra, G., & Klerkx, J. (2014). Learning Dashboards: An Overview and Future Research Opportunities. Personal and Ubiquitous Computing, 18(6), 1499-1514.

Zorilla et al. (2015)

Zorrilla, M., Á lvarez, E., & García-Saiz, D. (2015). A Parametrisable Method for Measuring Online Attendance in E-Learning Tools. International Journal of Technology Enhanced Learning, 7(4), 289-308.

Appendix A: Wordcloud LAK

Appendix B: Search String

Search String ProQuest (EDUCATION COLLECTION, ASSIA, IBSS, PSYCINFO) – 03.04.2016

(TI,AB("computer" OR "blended learning" OR "computer aided" OR "computer assisted assessment" OR "computer assisted learning" OR "computer based assessment" OR "computer based learning" OR "computer based teaching" OR "computer simulation" OR "computer supported" OR "computer technology" OR "computer use" OR "computer aided" OR "computer-assisted assessment" OR "computer assisted learning" OR "computer-based assessment" OR "computer-based learning" OR "computer-based teaching" OR "computerized instruction" OR "computers and learning" OR "computers in education" OR "computer-supported" OR "digital learning" OR "digital technology" OR "educational technology" OR "e-learning" OR "e-assessment" OR "electronic learning" OR "game" OR "ICT" OR "information communication technology" OR "innovative technology" OR "instructional technolog*" OR "intelligent tutoring system" OR "interactive learning environment" OR "interactive learning object" OR "interactive simulation*" OR "interactive white board*" OR "media in education" OR "mobile learning" OR "multimedia learning" OR "OLPC" OR "one laptop per child" OR "one to one computer" OR "one2one computer" OR "online learning" OR "online study" OR "simulation-based education" OR "simulation-based teaching" OR "simulation" NEAR "student" OR "simulation" NEAR "learn*" OR "tablet" OR "technology-enhanced education" OR "technology-enhanced assessment" OR "technology-enhanced learning" OR "technology use" OR "technology enhanced instruction" OR "technology enhanced assessment" OR "technology enhanced learning" OR "TEL" OR "tutoring system" OR "virtual learning" OR "virtual reality" OR "web-based instruction" OR "web-based learning" OR "web-based training" OR "CBAfL" OR "computer based Assessment for Learning" OR "computer-based Assessment for Learning" OR "learning analytics" OR "LA" OR "data driven" OR "educational data mining" OR "EDM" OR "tool" NEAR "student" OR "tool" NEAR "learn*" OR "game-based assessment" OR "game based assessment")) AND (TI,AB("formative assessment" OR "assessment for learning" OR "AfL" OR ("feedback" W/5 "student*") OR "peer assessment" OR "peer-assessment" OR "self-assessment" OR "assessment literacy")) NOT (TI,AB("higher education" OR "vocational education" OR "early years education" OR "pre school" OR "pre-school" OR "medic*" OR "nurse*" OR "health"))

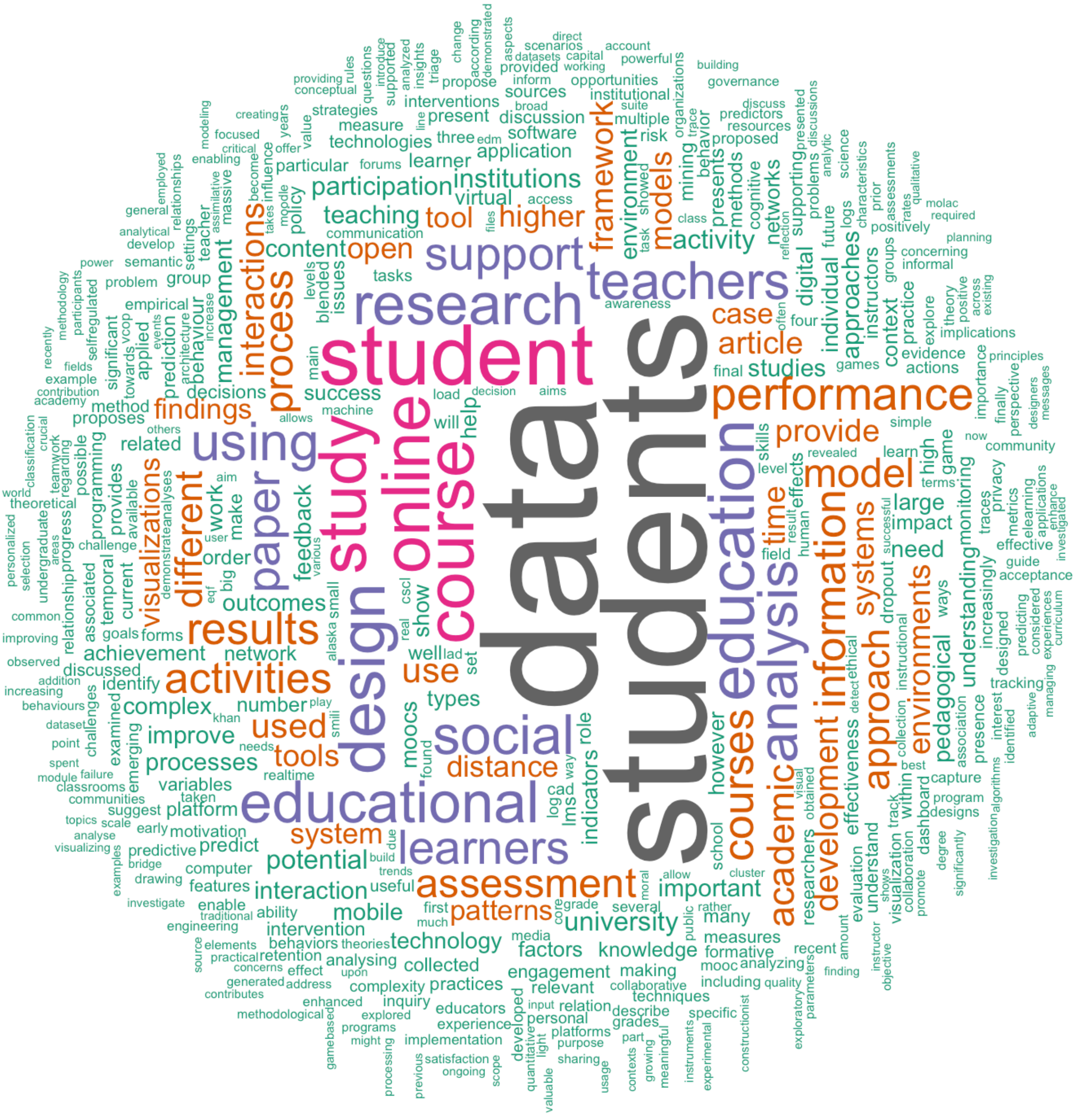
Filter: Peer-reviewed, published after 01.01.2012 Hits: 271

Search String SCOPUS – 03.04.2016

TITLE-ABS-KEY(“computer” OR “blended learning” OR “computer aided” OR “computer assisted assessment” OR “computer assisted learning” OR “computer based assessment” OR “computer based learning” OR “computer based teaching” OR “computer simulation” OR “computer supported” OR “computer technology” OR “computer use” OR “computer aided” OR “computer-assisted assessment” OR “computer assisted learning” OR “computer-based assessment” OR “computer-based learning” OR “computer-based teaching” OR “computerized instruction” OR “computers and learning” OR “computers in education” OR “computer-supported” OR “digital learning” OR “digital technology” OR “educational technology” OR “e-learning” OR “e-assessment” OR “electronic learning” OR “game” OR “ICT” OR “information communication technology” OR “innovative technology” OR “instructional technolog*” OR “intelligent tutoring system” OR “interactive learning environment” OR “interactive learning object” OR “interactive simulation*” OR “interactive white board*” OR “media in education” OR “mobile learning” OR “multimedia learning” OR “OLPC” OR “one laptop per child” OR “one to one computer” OR “one2one computer” OR “online learning” OR “online study” OR “simulation-based education” OR “simulation-based teaching” OR (“simulation” W/5 "student") OR ("simulation" W/5 "learn*") OR “tablet” OR “technology-enhanced education” OR “technology-enhanced assessment” OR “technology-enhanced learning” OR “technology use” OR “technology enhanced instruction” OR “technology enhanced assessment” OR “technology enhanced learning” OR “TEL” OR “tutoring system” OR “virtual learning” OR “virtual reality” OR “web-based instruction” OR “web-based learning” OR “web-based training” OR “CBAfL” OR “computer based Assessment for Learning” OR “computer-based Assessment for Learning” OR “learning analytics” OR “LA” OR “data driven” OR “educational data mining” OR “EDM” OR (“tool” W/ 5 "student") OR ("tool" W/5 "learn*") OR “game-based assessment” OR “game based assessment”) AND TITLE-ABS-KEY(“formative assessment” OR “assessment for learning” OR “AfL” OR (“feedback” W/5 "student*") OR “peer assessment” OR “peer-assessment” OR “self-assessment” OR “assessment literacy”) AND NOT TITLE-ABS-KEY(“higher education” OR “vocational education” OR “early years education” OR “pre school” OR "pre-school" OR "medic*" OR "nurs*" OR “health") (LIMIT-TO (PUBYEAR , 2017) OR LIMIT-TO (PUBYEAR , 2016) OR LIMIT-TO (PUBYEAR , 2015) OR LIMIT-TO (PUBYEAR , 2014) OR LIMIT-TO (PUBYEAR , 2013) OR LIMIT-TO (PUBYEAR , 2012)) AND (LIMIT-TO (DOCTYPE , "article") OR LIMIT-TO (DOCTYPE , "review") OR LIMIT-TO (DOCTYPE , "article in press")) AND (LIMIT-TO (LANGUAGE , "English"))

Hits: 1277

Appendix C: Wordclouds with and without learning analytics



data
students
student
research
teachers
support
online
course
study
design
educational
education
analysis
social
learners
using
paper
performance
model
information
academic
development
approach
courses
assessment
activities
results
process
interactions
framework
environments
systems
different
visualizations
patterns
distance
tools
used
use
system
higher
tool
open
content
teaching
participation
institutions
models
activity
case
article
studies
provide
findings
feedback
interaction
potential
university
important
technology
mobile
processes
improve
complex
achievement
network
outcomes
success
decisions
management
environment
networks
pedagogical
understanding

SLATE

SlateResearch SlateResearch www.slate.uib.no

UNIVERSITY OF BERGEN