Learning Analytics: Using Data-Informed Decision-Making to Improve Teaching and Learning: Maximizing Student Engagement, Motivation, and Learning

Chapter · January 2019

DOI: 10.1007/978-3-319-89680-9_7

Alyssa Friend Wise
New York University

PRE-PRINT COPY

To appear in: O. Adesope & A.G. Rudd (Eds.) *Contemporary technologies in education: maximizing student engagement, motivation, and learning*. New York: Palgrave Macmillan.

Learning Analytics:
Using Data-Informed Decision-Making to Improve Teaching and Learning

Alyssa Friend Wise, *New York University*

Abstract

Learning Analytics is the development and application of data science methods to the distinct characteristics, needs, and concerns of educational contexts and the data streams they generate for the purpose of better understanding and supporting learning processes and outcomes. This chapter describes three characteristics of Learning Analytics work that distinguish it from prior educational research to give readers a concise overview of what makes learning analytics a unique and especially promising technology to improve teaching and learning. *Data* used in learning analytics relates to the process of learning, can come from a variety of sources (in both virtual and physical learning environments), and are characterized by their large quantity and relatively small grain-size. The most common forms of data are activity, artifact and interaction information. *Analysis* approaches aim at detecting underlying patterns and relationships in the data and include prediction, structure discovery, temporal, language-based and visual methods. *Pedagogical Uses* are what position learning analytics as more than simply a new set of methods but an impactful technology to drive data-informed decision-making through tailoring educational experiences, informing student self-direction, and supporting instructor planning and orchestration. The chapter concludes with an overview of the systemic and societal issues surrounding learning analytics use that will frame how and to what extent they are able to affect education.

Introduction

Learning Analytics is the development and application of data science methods to the distinct characteristics, needs, and concerns of educational contexts and the data streams they generate for the purpose of better understanding and supporting learning processes and outcomes (see also an earlier definition by Siemens et al., 2011). It is both a field of scholarly pursuit and a technology for making concrete improvements within educational systems by enabling data-informed decision-making by teachers, students, and other educational stakeholders. Learning Analytics has been identified as a critical emerging technology of the 21st century, with high expectations to make a positive impact on learning and teaching (Johnson et al., 2016), both through short-cycles improvements to educational practice and long-cycle improvements to our understanding of learning.

While the collection and analysis of data to understand and support learning is not a new endeavor, there are three critical characteristics that distinguish learning analytics from prior educational research: the data the work is based on, the kinds of analyses employed and the ways in which it is put to use. This chapter begins with an overview of the value proposition that learning analytics offer and is then organized around these three areas (Data, Analyses, and Applications) to give readers a concise overview of what makes learning analytics a unique and especially promising technology to improve teaching and learning.

Why Develop Learning Analytics?

The basic value proposition for learning analytics is that generating more information about how learning processes unfold can help us better improve them. This is true over the long term (by better understanding how learning occurred in the current situation we can design in a more informed way for the future), but also in the short term (by better understanding how learning is occurring up to the current moment, we can act in a more informed way right now). It is the latter use, to improve teaching and learning in "real-time" that is most novel and exciting to educators. From an instructor's perspective, learning analytics can both provide a way to check if the planned activities are occurring as intended (e.g. the goal is for pairs of students to argue opposing positions about the culpability of Lady Macbeth – is this actually happening?) and to identify particular groups that may need additional support (e.g. in some groups the conversation is balanced but in others one student dominates over their partner). Similar information can also be provided directly to students (either individually or in collaborative groups) to prompt reflection and regulation of their own learning processes. Another attractive use of learning analytics is to tailor educational experiences to better meet the specific needs of one of more students. Higher education has been accused of a "one size fits all" approach, in part because identifying meaningful difference between students' needs and acting on them each appropriately is incredibly time consuming when done manually by instructors. However, if the right dimensions of difference are known and can be detected in naturally generated data, then the vision of education tailored to each students' needs becomes tractable.

What Kinds of Data are Learning Analytics Based On and What Makes Them Distinct?

Learning analytics is not defined primarily by the source of data but by its size. Size here refers to two distinct characteristics. The first is the *overall quantity of data* involved. Simply put, the computational analyses used in learning analytics generally require a greater amount of data than that used in traditional educational research. The larger amount can, in part, come from a greater number of people; however, in larger degree it is a result of collecting a much greater number of measurements on each person. So, for example, learning data from MOOCs (massive open online courses) is large not just because there are thousands of learners, but because we can collect a data point for every single action a person takes in the system (amounting to at least tens of thousands of data points per learner). The second element of size is the *granularity of the individual data points* themselves. Here, the measurements taken are generally more micro than traditional data, with learning analytics often looking at fine-grained elements of the learning process. Importantly, the smaller grain size is not created artificially to inflate the data available (for example, taking the temperature of a room every 30 seconds instead of every 30 minutes creates more data but not necessarily more information). It is a reflection of new tools that allow for the capture of learning activity at the grain size which it actually occurs; i.e. action by action. Together, smaller and more numerous data points are a hallmark of learning analytics research.

Source, Quantity and Granularity of Learning Analytics Data

Where does the size of learning analytics data come from? The increase in availability of large quantity / small unit size data can be attributed, at least initially, to the rise of digital technologies used for learning. From generic learning management systems (LMSs) to focused intelligent tutoring systems (ITSs), from virtual discussion boards (and other social media tools) to face-to-face classroom response systems (clickers) the dramatic rise of technologies used to support teaching and learning has facilitated the efficient collection of diverse (though not comprehensive) forms of data from large numbers of students at many points in time. This aligns with the essential attributes of big data, described as volume, velocity, and variety (Laney, 2001)

For example, while previously an instructor might record the overall grades of class members on a quiz, online tools can easily track item-level responses for everyone in the class on every assessment across the term. Similarly, once built, technologies lend themselves to use at scale allowing the collection of data from much larger numbers of students than was possible previously (this is true for both formal learning environments such as MOOCs as well as informal learning support tools such as Piazza or even Twitter). Furthermore, while prior data were limited to what could be captured in the classroom (or self-reported by students), internet based tools allow (potential) insight into student learning regardless of where it takes place. In short, as more and more of our (academic) lives take place with the support of digital tools, the virtual “footprints” we leave behind also become more detailed and abundant.

Kinds of Learning Analytics Data

Learning analytics data relates to the *process* of learning (as opposed to just its outcomes). Current forms of data commonly used in learning analytics work include activity data (traces of what students did) and artifact data (things that students created). Often, a single action by a student can produce both kinds of data: for example, if a student attempts a quiz question in an online tool there is an activity trace (student X answered question Y at time T) and an artifact created (the actual answer they gave, which might later be evaluated in some way). A third form, association data, is often constructed based on the prior two to index relationships between students and students, students and artifacts, or students and instructors (see Hoppe’s description of a trinity of learning analytics approaches aligned with these data types in Suthers et al., 2015). In addition to these core data sources about the learning process, learning analytics may also incorporate other kinds of data, such as learning outcomes (either prior or current performance) and demographic information (Sclater, 2017); since these data are pre-existing rather than generated during the course of learning, they can be considered in the category of archival data. Traditional self-reports are less commonly used in learning analytics research due to problems of inaccurate and selective recall related to learning behaviors and the degree of intrusion required to collect the data (Baker & Siemens, 2014; Winne, 2010). However, if there is a need to document aspects of students' perceptions, experience sampling methods (ESM, Csikszentmihalyi & Larson, 2014) via mobile apps or the like can be used. Finally, learning environment data (such as a course's curriculum or pedagogical approach) can be important as an element of meta-data (or secondary level data in a hierarchical analysis approach if multiple courses are studied) to contextualize the primary-data and determine appropriate approaches to analysis.

Activity data most commonly take the form of log-file data, a record of actions the student took in an online system at specific points in times. Log-file data can be coarse or detailed depending on both the front-end user interface and back-end data structure. For example, an LMS record may indicate that a student "opened message #241783 in discussion #486" or simply that they "accessed the discussion forums." Similarly, some systems capture the exhaustive use of play/pause/rewind/ fast-forward controls used during video-playback, while others only indicate that a video was viewed. In addition to LMSs, activity data can also come from the use of digital library resources, e-books (if the publisher provides access), and other dedicated learning tools housed outside the LMS (e.g. adaptive testing, intelligent tutors, simulations). While more instrumentation is required, activity data can also be collected from physical learning environments via multi-modal learning analytics tools. Multi-modal data can include the tracking of student gaze, gesture, posture, movement as well as physiological measures such as heart rate, galvanic skin response, and electroencephalogram (EEG) readings (Ochoa & Worsley, 2016).

Artifact data can be any object created by a student and stored by the system. Here, the level of granularity corresponds to the unit-size submitted by the student. While audio, image, and video artifacts are certainly possible (e.g. Baltrušaitis, Robinson, & Morency, 2016; D'Angelo, Roschelle, & Bratt, 2015), by far the most common type of artifacts are text-based, and include objects such as answers to questions, discussion forum posts, student essays and even lines of code. Artifact data must undergo some assessment or decomposition during the analysis process to index a number of its qualities. In simple cases, an artifact such as a question answer (e.g. the number 7 entered by a student in response to the question "3x2=?") might be evaluated as correct or incorrect. In more complex cases, a series of metrics might be used to represent the artifact. For example, a student essay could be indexed by its word length, structural coherence (McNamara, Crossley, & McCarthy, 2010), and the extent to which vocabulary from the course readings was employed (Velazquez, Ratté, & de Jong, 2016). While traditional teacher assessment and educational research often involves the evaluation of student work manually by human raters using a rubric (for example in terms of the quality of writing, strength of evidence presented, or justification of positions taken), learning analytics requires such evaluation to occur at scale. Thus, one major stream of learning analytics research is devoted to the development of computer models that can "learn" to perform this task based on a training set of human-coded data (Mu et al., 2012). Artifacts can also be used to infer qualities of the student producing them, for example in the intelligent tutoring system (ITS) literature where students' answers to problem-solving steps are used to build a model of the students' underlying knowledge state (Corbett & Anderson, 1994).

Association data are generally constructed post-hoc from activity and/or artifact data. Associations can be made on the basis of similarity (e.g. two students took the same course) or interaction (e.g. one student sent another student (or the instructor) a message). Associations can also exist between people and artifacts (e.g. a student accessed a certain video resource) or between two artifacts (e.g. similar vocabulary used across two essays). When using association data, it is important to be clear about what kind(s) of elements are being associated and what the nature of the association (similarity, interaction etc.) indicates. The existence of association data points to an important question in learning analytics about the unit of the analysis. Most learning analytics to date have focused on the individual student (and their activity, their artifacts, their associations) as the object of interest. However, there is increasing interest in collaborative learning analytics in which the object of interest is a small group or community (e.g. Chen & Zhang, 2016).

Data, Features, and Proxy Indicators

More data do not necessarily mean more information and an important challenge that learning analytics work must address is crafting meaningful indicators from what is available. Because learning analytics researchers do not always have control over the design (both front-end interface and back-end data structure) of the tools from which they collect data, they must be creative in devising proxies, measurements that serve as reasonable representations of the construct or phenomenon they wish to study. From an educational perspective, the justified linking of an observation to a conceptual entity is a critical piece of the logic chain for establishing the validity of learning analytics work. For example, should more time spent in an LMS be taken as an indicator of engagement or effort? The answer may depend on if the time relates to solving a problem (more time indicates the student exerted more effort) or reading discussion posts (more time indicates more engagement). Aggregating the overall time is problematic as it confounds the two.

From a data science perspective, the problem of selecting indicators focuses on how to transform the raw data into a set of features that best models the underlying phenomena. This process is referred to as feature engineering and includes feature construction (e.g. via various forms of aggregation or decomposition), feature extraction (e.g. via dimensionality reduction techniques such as principle component analysis), and feature selection (choosing a subset of possible features to include based on some ranking of their anticipated importance in the model). See Sinha, Jermann, Li & Dillenbourg (2014) for a particularly nice example of engineering interpretable learning features from low-level data using fuzzy pattern matching.

It is a debated question in the field as to what extent it is important for engineered features to be interpretable versus simply contribute strongly to a prediction (see Bergner, 2017, pp 41-42 for elaboration of the differences between explanatory and predictive models). While there are some cases where reliable prediction alone is useful, in the realm of education we generally want to understand why certain relationships exist and be able to take action to affect them. For example, it is difficult to help a student who is identified as being at-risk for failing a course if there is no way to make sense of the factors that led them to be placed into this category. There are also concerns with the use of features that might (unintentionally) reinforce traditional educational inequalities (Slade & Prinsloo, 2013). For these reasons, theory can be a powerful tool to constrain and shape the possible degrees of freedom for constructing, extracting, and selecting features (Wise & Shaffer, 2015).

The Manufacture of Data

Finally, it is important to remember that learning data are neither natural nor neutral. Learning data are not "natural" because they are produced as students interact with designed environments. The data thus represent aspects of what students do in response to that *specific environment*. In order to properly generalize to other contexts, we need to index the important qualities of the environment which range from the technical (e.g. the tools available, how they are designed, interface and navigational features) to the pedagogical (e.g. is the course oriented towards acquisition of facts, problem solving skills, construction of conceptual schema). Learning data are also not "neutral" in that what is captured is often as much a product of what is feasible as what is valuable. The data that are easiest to acquire may not be the most useful or important; for example indexing students' activity based solely on their use of an LMS when in-class lectures and tutorials are a greater part of the course's pedagogy may produce a skewed picture. Furthermore, once LMS data are reified as a measure of "student activity," they becomes a target to be optimized. Thus students who are active in class and tutorials, but less so online, may feel misplaced pressure to increase their use of the LMS. In general it is easier to try to improve one's standing on metrics that do exist, than to remember the value of those things which we can't (yet) quantify; thus we run the danger of becoming what we measure (Duval & Verbert, 2012). As the field of learning analytics matures we expect to see learning tools for which the design of the data produced is an integral concern from the start rather than an afterthought. This will generate more useful data both through better back-end structures and through the creation of front-end interfaces that more readily support inference-making from data.

What Kinds of Analyses Does Learning Analytics Employ and What Can They Tell Us?

Learning analytics methods includes the set of human and computational processes and tools used to manipulate data in order to produce meaningful insight into learning. Much learning analytics works draws on educational data mining approaches (see Romero, Ventura, Pechenizkiy, & Baker, 2010), though given that learning analytics also seeks to attend to underlying conceptual relationships and the situational context, the metaphors of data geology and data archeology have been proposed as more appropriate than that of mining (Wise & Shaffer, 2015). Avoiding the politics of language, learning analytics can be said to employ educational data science methods to detect underlying relationships and patterns among variables and cases. There are several classes of methods commonly used to achieve this. Each is discussed below with an emphasis on applications, i.e. the kinds of things that can be learned from each approach and the ways it can be used to support learning. In line with this focus on application, the references provided offer examples of the ways each approach has been employed to provide insight into educational data, rather than serving as authoritative sources on the technical details of the method.

PRE-PRINT COPY

To appear in: O. Adesope & A.G. Rudd (Eds.) *Contemporary technologies in education: maximizing student engagement, motivation, and learning*. New York: Palgrave Macmillan.

Prediction (Supervised) Approaches

One of the most common and useful approaches in learning analytics is prediction (Baker & Yacef, 2009; Papamitsiou & Economides, 2014). Prediction is a form of supervised machine learning; the “supervision” refers to the fact that values for the thing being predicted (the target) are known a priori for a training / test data set and thus the accuracy of the model can be evaluated with respect to these known values. Prediction models use a combination of attributes for a case (the predictor variables) to predict the value of another attribute (the target).

Prediction models can produce several different kinds of results useful to learning analytics. First, they can be used to forecast an attribute for a case (e.g. an assessment score or at risk status for a student) when it is not known, either because it was not collected or it is related to something that has not yet occurred. A common application of this is early-alert systems developed by universities to identify students at risk for poor performance or dropping out (Arnold, 2010). For example, Jayaprakash et al. (2014) developed a classifier that predicted whether students were likely to earn a grade of C or higher in a course (“successful completion”) or not (“unsuccessful”). Their model was build based on a combination of attributes including demographics and academic records (archival data), prior scores (evaluated artifact data) and LMS usage (activity data). With a predictive goal in mind, Jayaprakash et al. (2014) were interested in developing an accurate model so that they could apply it to students at the start of the course to forecast which students were likely to be unsuccessful based on the academic records, prior scores and LMS data. When the predicted value is correctness on future learning assessments the result is often used to drive adaptation in systems such as intelligent tutors (see Corbett & Anderson, 1994 for an expanded explanation of knowledge tracing). A special case of forecasting that is particularly useful in learning analytics is the combination of prediction models with Natural Language Processing techniques (see description below) to perform automated or semi-automated content analysis of artifact data (Cui, Jin & Wise 2017; Rosé et al., 2008). This can be used to provide feedback to students or instructors on the work performed.

A second kind of use of prediction models is explanatory. In this case the focus is not on forecasting values for new students but to better understand relationships between variables (though unless factors were manipulated experimentally, claims of causality should be avoided). For example, Svihla, West and Linn (2015) used six different log-file metrics indexing the different ways students (the cases) revisited content in an online inquiry-learning tool (activity data predictors) to predict their score on a delayed cumulative project assessment (evaluated artifact data target). Their results showed that distributed visitation of a dynamic visualization was predictive of students’ understanding of the content several weeks after the unit had been completed. Svihla et al.’s (2015) explanatory use of their model it allowed them to make claims about the relationship between distributed revisiting and maintenance of understanding over time.

When prediction targets continuous variables (e.g. the delayed assessment score in Svihla et al., (2015), models such as linear regression, support vector machines and regression trees are commonly used. For categorical (including binary) outcome variables, classification models (aka classifiers) are built. Common classification methods include decision trees, logistic regression, naïve bayes and support vector machines. The quality of prediction models can be evaluated in various ways such as calculating accuracy, precision-recall values, AUC or other metrics (see Zheng, 2015); these metrics should be reported on cross-validation and external test-sets, not the same training set on which the model was developed. Generalizability can be assessed using similar metrics on external test-sets from different learning contexts (e.g. different populations, different years, different subject matter, different pedagogy). There is an inherent tradeoff in building models: sensitivity to specific features of a learning context comes at the cost of broad applicability to multiple situations while models built to be used across a wide range of contexts will be less sensitive to the data available in any particular one (Gašević et al., 2016).

Structure Discovery (Unsupervised) Approaches

Structure discovery is another common analytic approach that offers different ways to find patterns of similarity or relationship among cases (e.g. students, messages, essays, curriculum) or variables (attributes of the cases). Unlike prediction, there is no predefined target to model or evaluate success against (for this reason, structure discovery methods are called unsupervised machine learning). Structure discovery methods such as correlation mining, association rule mining and factor analysis are useful to identify regularities in *variables* (e.g. students who rewind and re-watch online videos more, tend to ask questions in the discussion forum; multiple variable seem to index different facets of engagement in the course). Structure discovery methods such as clustering, social network analysis, and topic modeling are generally used to identify commonalities and differences between *cases* (e.g. these resources are used by students early, but not late, in a course; these student tend to access many resources but do poorly on quizzes).

Correlation and association rule mining are similar to prediction in that the underlying algorithms identify recurring relationships between variables; however, relationships may be found between *any* combinations of variables. Correlation mining focuses on linear relationships between continuous variables (e.g. the more time a student spends on online practice questions the higher their grade on the actual test) while association rule mining is typically used to generate if-then rules about the co-occurrence of categorical variables (e.g. if a student takes both Biology and Chemistry they are likely to also take Biochemistry). Given the large number of possible variables and relationships that may be identified due to chance, it is important to carefully control for false discovery (see Hero & Rajaratnam, 2016) and to critically evaluate results with respect to both empirical standards (e.g. see discussion of measures of support, confidence, and interestingness by Merceron & Yacef, 2008) and theoretical soundness (Wise & Shaffer, 2015).

Factor analysis is a technique that finds groups of continuous variables whose values (for a given population) consistently align with each other and thus can be combined as a representation of some latent factor. This can both provide insight into the underlying structure of constructs that the variables index and also be used for dimensionality reduction. Dimensionality reduction (which can be also be achieved by Principal

Component Analysis) is important to avoid over-fit and uninterpretable models. For example, Ahn (2013) used factor analysis to reduce 12 variables of Facebook usage data collected from university students (e.g. wall posts made, links shared) into four latent factors representing different classes of Facebook activity: messaging, information sharing, friending and affiliating. The factors were then input into a regression model to predict the students' new media literacy skills.

Clustering, social network analysis, and topic modelling differ from the above methods in that the focus is on generally on regularities in *cases* rather than *variables*. Clustering is commonly used to identify cases (often students, but at other times resources, courses etc.) who consistently have similar values to each other across multiple variables, and thus can be thought of as being of the same "type." For example, Wise, Speer, Marbouti and Hsiao (2013) performed a cluster analysis on log-file data indexing how students "listened" and "spoke" in online discussions to identify three underlying groups: *Superficial Listeners, Intermittent Talkers*; *Concentrated Listeners, Integrated Talkers*; and *Broad Listeners, Reflective Talkers*. Importantly, as labeling clusters is a task of human interpretation, it can be useful to look closely at the data: Wise et al. performed targeted case studies on a representative member of each cluster that contributed important insight to cluster labels beyond that available from the aggregate variable values.

Topic modelling is a form of text-mining (other text-mining techniques are discussed below) that is used to represent the underlying structure across a corpus of documents (which could be student essays, social media messages, etc.) by identifying collections of topics (sets of co-occurring words) and the extent to which they are present in each document. A common application of topic modeling is to make sense of the large volume of messages that are contributed to online course discussions, MOOC forums, and social media. For example, Joksimović et al. (2015) examined what MOOC participants talked about in various social media venues and Vytasek, Wise, & Woloshen (2017) explored how topic models could provide classroom instructors with a useful big picture view of large and diverse online discussions.

Finally, social network analysis (SNA) is a technique that looks for regularities not in the attributes of the cases themselves but in the relationships between them. This can provide insights about individuals (e.g. measures of their centrality), the entire network (e.g. its density), or some sub-set of it (e.g. the presence of cliques). A key decision in SNA is how to define the nodes and the linkages between them. A common approach is to take nodes as students and to create linkages based on their interaction (e.g. Wise, Cui, & Jin, 2017); however linkages based on similarities (e.g. Hecking, Chounta, & Hoppe, 2016) and bi-partite networks which include both individuals and objects they interact with (e.g. resources accessed) are also possible (Poquet & Dawson, 2016). SNA has been useful in understanding about general characteristics of social interactions and relationships (e.g. Dowell et al., 2015), exploring their relationship with learning outcomes (Dawson 2010; Rabbany, Takaffoli, & Zaïane, 2011), and identifying small groups within larger networks worthy of more detailed attention (Wise, Cui & Jin, 2017). While standard SNA approaches produce descriptions of connections in aggregate, more sophisticated techniques such as

ERGM (exponential random graph models) and dSNA (dynamic social network analysis) allow for inference testing and the study of network evolution over time respectively (e.g. Joksimović et al., 2016; Zhu et al., 2016).

Temporal Approaches

Similar to structure discovery, temporal approaches to data analysis also look to discover previously undefined patterns in the data, but in this case, the patterns relate to the sequence and flow of events over time (Knight, Wise, Chen, & Cheng, 2015). Temporal approaches are a particularly important set of methods in learning analytics as they leverage traces of activity to address the field's fundamental concern with studying and understanding learning as a process (Suthers et al., 2015); however, they have been underutilized in the field of learning analytics thus far (Chen et al., 2016). Temporal approaches in learning analytics can be roughly divided as those which deal with time explicitly through examination of flow and fluctuation in features of the learning process over time (e.g. survival analysis, Yang, Sinha, Adamson, & Rose, 2013) and those which deal with time implicitly through examination of sequences of events in the learning process (e.g. lag sequence analysis, Chen & Resendes, 2014; (hidden) markov modelling, Jeong, Biswas, Johnson, & Howard, 2010). Temporal analyses can also be used to divide a learning process up into different phases of activity (e.g. via sequential discourse analysis, Wise & Chiu, 2011).

Natural Language Processing Approaches

Natural language processing (NLP) approaches in learning analytics use computational techniques to assess various linguistic features of texts (McNamara et al., 2017). It is an exciting area in active development that allows for direct inspection of a wide variety of textual data sources that includes: standalone student artifacts such as student essays or short answers; traces of dialogue among students and instructors; and collections of instructional resources. Roughly these differences align with the distinct concerns and applications of writing analytics (Shum et al, 2016), discourse analytics (Knight & Littleton, 2015; Rosé, 2017), and content analytics (Kovanović et al., 2017). NLP approaches are frequently used in combination with other analysis approaches already discussed including prediction (e.g. Mu et al., 2012), structure discovery methods (e.g. Dowell et al, 2015) and temporal analysis (e.g. Suthers & Desiato, 2012). NLP approaches useful for learning analytics extract linguistic features about words and their assemblages. Analyses performed on words may assess basic presence (e.g. frequency of particular n-grams, parts of speech or LIWC (Linguistic Inquiry Word Count) categories) or delve more deeply into their underlying meaning (e.g. via LSA (latent semantic analysis, different from the temporal technique of lag-sequence analysis, see Landauer, MacNamara, Dennis & Kintsch, 2011 for a wide-ranging overview of theory, methods, and applications of the technique). Other techniques examine the use of particular parts-of-speech (such as verbs), syntactic structure, and the cohesion across a text (McNamara et al., 2017). When considering relations between texts measures of semantic similarity (often calculated using LSA) are particularly useful.

Visual Approaches

Much of the work in learning analytics using visualization is not actual visual analysis per se, but the visualization of the outputs of other analyses for communication with various stakeholders. Learning analytics dashboards, for example, often employ static graphical representations of analytic results designed to evoke particular responsive actions (Klerkx, Verbert & Duval, 2017). In contrast, true visual analytics exploit visualization techniques and human perceptual abilities as part of the analytic process itself (Shneiderman, 2014). This is done by visually representing data in ways that support human recognition of patterns and aberrations, often via an interactive interface that allows for manipulation and permutation of the visualizations (Ritsos & Roberts, 2014). While some limited examples of static visual learning analytics exist, for example human inspection of heat maps (Pecaric, Boutis, Beckstead, & Pusic, 2017; Serrano-Laguna, Torrente, Moreno-Ger, & Fernández-Manjón, 2014) and moment-by-moment learning curves (Baker et al., 2013), there is great room for further development of interactive visual analytics.

What Kinds of Pedagogical Uses Can Learning Analytics Serve and How Do They Support Learning?

Tailoring Educational Experiences

An initial class of pedagogical use of learning analytics is for tailoring educational experiences to better meet the specific needs of one of more students. In this model of use, the analytics are used to create some sort of a (static or dynamic) profile of learners with the educational experience provided for them differing in response to this. This has been referred to at times under the label of "personalized learning," but such terminology is overly narrow because it assumes that the target for the tailoring is an individual, when it could also be a group of learners, and it implies that the activity is done for learners, when in many cases the learner must actively take up a recommendation that is provided. Tailoring of educational experiences for individuals or groups can occur through both adaptive (computer driven) and adaptable (human driven) changes to a system that make it more appropriate for the learning of those involved. Common analytic techniques that drive tailoring include prediction models, clustering to identify groups of students with similar profiles, and association rule mining.

A high-profile class of tailoring applications are adaptive systems in which the resources, questions or other learning materials provided to students are determined based on an underlying analytics model. One of the earliest set of adaptive learning tools were intelligent tutoring systems which construct a model of both the domain and learner in order to provide immediate customized feedback to students (Nwana, 1990). Recently a large number of companies, including textbook publishers, have also moved into the adaptive learning space. Adaptive learning tools may be designed around specific pre-determined content or exist as platforms for instructors or institutions to input their own content. They do not need to only involve static content and problems to be solved but can incorporate (or be embedded in) games and simulations as well. It is important to distinguish between tools which make adaptations directly based on learner activity versus those which use more sophisticated approximations of learner's cognitive skills.

Different from adaptive tools in which the tailoring is fully enacted by a system and may not be apparent to the learner, recommendation engines are systems that provide tailored suggestions (of courses or learning resources) to students. Two well-known course recommendation systems are Stanford's CourseRank system (Parameswaran, Venetis, & Garcia-Molina, 2011) and Degree Compass (Denley, 2013), originally developed at Austin Peay State University, and recently acquired by Desire to Learn. Systems for recommending useful learning resources (or useful sequences of resources) are generally developed in the context of particular learning tools (see Drachsler et al., 2015 for a review of 82 different recommender systems). Finally, early-alert systems use predictive models to identify students at-risk of failing a course or dropping out of university. Studies have shown that simply making students aware that they are at-risk can have an impact on their academic standing (Arnold, 2010); though providing students with actionable strategies is much preferred. In a recent review of early alert systems, Sclater (2017) points out that more evidence about when and why such systems are effective is needed.

Informing Student Self-Direction

Different from tailoring the materials that are given to students, another pedagogical use of analytics is to support students in conscious attention to and improvement of their own learning processes. This model of use draws heavily on psychological theories of experiential learning Kolb (1984), self-reflection Schön (1983), and self-regulated learning (Winne, in press) in which learning analytics provide feedback that students can use to adjust or experiment with changes in their learning behaviors. A wide variety of tools exist to provide students with feedback on their academic status and study habits (e.g. E2Coach at the University of Michigan, Huberth et al., 2015; Check My Activity at the University of Maryland, Baltimore County, Fritz, 2011), essays (e.g. OpenEssayist, Whitelock, Twiner, Richardson, Field, & Pulman, 2015), and discussion forum participation (e.g. E-Listening Analytics Suite, see Wise, Zhao, & Hausknecht, 2014). Such feedback can be provided in a variety of forms which may be embedded directly into the learning environment or extracted from it (Wise et al., 2014), for example via email messages or real-time dashboards that can be accessed at any time. The challenges for students in interpreting and using such information are great, however, and the most powerful systems provide not only the analytic feedback but also some sort of structure or support for making sense of and acting on the information provided (Wise & Vytasek, 2017).

Supporting Instructor Planning and Orchestration

For instructors, pedagogical uses of learning analytics can be used to support refinement of both the overarching learning design and the decisions they make to orchestrate classroom activity within it. From the perspective of learning analytics and learning design, analytics offer a way to empirically verify (or refute) assumptions about the classroom (be it physical or virtual). The process for doing so requires instructors to document their pedagogical intentions (the design), describe activity patterns that indicate fulfillment of these intentions (targets), and then use the analytics to evaluate the degree to which the patterns occurred (Lockyer, Heathcote & Dawson, 2013). Systems that provide feedback to instructors about their learning design are typically presented via teacher dashboards (see review in Verbert, Duval, Klerkx, Govaerts, & Santos, 2013). Examples of this cycle in action are given in Brooks, Greer, and Gutwin (2014) who look at instructors' modification of their discussion forum practices based on SNA diagrams and Roll et al. (2016) who examined how course designers of a MOOC planned for revisions based on the analytic feedback provided to them.

In addition to supporting critical attention to the activity outcomes of course design, learning analytics can also assist instructors in orchestrating their class. Analytics can provide information that helps instructors identify struggling students (and ideally know how or why they are struggling), recognize groups that are collaborating more or less productively (van Leeuwen, 2015) and pinpoint prevalent points of difficulty for a class (Ali, Hatala, Gašević, & Jovanović, 2012). Ideally, the analytics are used not only to identify "problem" situations, but as part of a regular feedback mechanism of tuning and adjustment (Wise, Vytasek, Hausknecht, & Zhao, 2016). In addition, another way in which analytics can help inform orchestration is by identifying types of students (or student behaviors) that occur repeatedly. Such information can be used by instructors to more easily identify and address common patterns or can be fed back to create accommodations or greater support structures in the learning design.

What Are Key Issues for the Future of Learning Analytics?

The optimistic vision of learning analytics in higher education described above is far from inevitable. Others have countered such images of a rosy future with the potential for (intentional or unintentional) misuse of analytics leading to a dystopian future of oversight and control (Rummel, Walker & Aleven, 2016). There is also the concern that, like so many promising educational technologies, learning analytics will not live up to the hype and fail to make a substantial impact (Cuban, 2001; Ertmer, 1999). In a comprehensive review of the empirical research on learning analytics use to date, Ferguson et al. (2016) emphasize that expectations are yet to be realized and evidence of successful and impactful implementation is still scarce. Key systemic and societal issues that will determine the fate of learning analytics include deliberate consideration of the policy needs required to govern the ethical dimensions of analytics use and proactive planning for the required infrastructure.

In terms of infrastructure, universities need to consider now what kinds of data streams and stores they will want to be able to access in the next 5-10 years. Data infrastructure planning includes attention not only to what data will be collected, but how (and where) the data will be stored, what meta-data will be used to index the data, and how (and by whom) the data will be queriable. Critically, system interoperability and the integration of multiple data streams (e.g. from learning management systems, student information systems, external tools, and human input) are core technical challenges to be addressed. Going further, universities will need to think about the analytic literacy of those who will want to ask questions of the data and what tools, people, and processes are needed to support these activities for both research and day-to-day teaching and learning purposes.

In terms of policy, institutions need to put in place clear guidelines for practices around data and analytics use (Prinsloo & Slade, 2013). Specifically, policies are needed to: allocate responsibility for data assets and analytic processes; establish procedures for giving consent / opting-out of data use, providing students with access to their own data, and protecting student privacy; set-up systems to check that inferences made based on data and algorithms are valid and transparent; and maximize positive analytics implementations while minimizing any potential adverse impacts (Sclater, 2014). Importantly, as students are a critical stakeholder (and the primary intended beneficiary) of learning analytics, they should be consulted as such policies are developed (Slade & Prinsloo, 2014). Other overarching important ethical issues to keep in mind include broad attention to algorithmic accountability (ACM US Public Policy Council, 2017), maintaining institutional value on those things that are not well-indexed by analytics, and remaining vigilant for unintended systemic consequences.

Conclusion

Learning Analytics is the development and application of data science methods to the distinct characteristics, needs, and concerns of educational contexts and the data streams they generate. The goal is to better understand and support learning processes and outcomes through both short-cycles improvements to educational practice and long-cycle improvements to the underlying knowledge base. This chapter has overviewed the distinct character of the data used in learning analytics, the kinds of analyses applied, and the pedagogical uses to which the analytics are put; together these characteristics highlight why learning analytics is seen as an especially promising technology to improve teaching and learning. To make this vision a reality, universities will need to be proactive in building up the requisite technical and policy infrastructure.

References

ACM US Public Policy Council (2017). *Statement on algorithmic transparency and accountability.* Washington, DC: ACM.

Ahn, J. (2013). What can we learn from Facebook activity?: Using social learning analytics to observe new media literacy skills. *Proceedings of the Third International Conference on Learning Analytics & Knowledge* (pp. 135-144). Leuven, Belgium: ACM.

Ali, L., Hatala, M., Gašević, D., & Jovanović, J. (2012). A qualitative evaluation of evolution of a learning analytics tool. *Computers & Education, 58*(1), 470–489.

Arnold, K. E. (2010). Signals: Applying academic analytics. *Educause Quarterly*, *33*(1).

Baker, R. S., Hershkovitz, A., Rossi, L. M., Goldstein, A. B., & Gowda, S. M. (2013). Predicting robust learning with the visual form of the moment-by-moment learning curve. *Journal of the Learning Sciences*, *22*(4), 639-666.

Baker, R., & Siemens, G. (2014). Educational data mining and learning analytics. In K. Sawyer (Ed.), *Cambridge Handbook of the Learning Sciences* (2nd ed., pp. 253–274). Cambridge, MA: Cambridge University Press.

Baker, R.S. & Yacef, K. (2009) The state of educational data mining in 2009: A review and future visions. *Journal of Educational Data Mining*, *1*(1), 3-17.

Baltrušaitis, T., Robinson, P., & Morency, L. P. (2016). Openface: An open source facial behavior analysis toolkit. *Proceedings of 2016 IEEE Winter Conference on Applications of Computer Vision* (pp. 1-10). IEEE.

Bergner, Y. (2017). Measurement and its uses in learning analytics. . In C. Lang, G. Siemens, A. Wise & D. Gašević (Eds.) *Handbook of learning analytics* (1st ed, pp. 35-48). Edmonton, AB: SoLAR.

Brooks, C., Greer, J., & Gutwin, C. (2014). The data-assisted approach to building intelligent technology-enhanced learning environments. In Larusson, J. A. & White, B. (Eds), *Learning Analytics* (pp. 123-156). Springer New York.

Chen, B., & Resendes, M. (2014). Uncovering what matters: Analyzing transitional relations among contribution types in knowledge-building discourse. *Proceedings of the Fourth International Conference on Learning Analytics & Knowledge* (pp. 226-230). Indianapolis, IN: ACM.

Chen, B., Wise, A. F., Knight, S., & Cheng, B. H. (2016). Putting temporal analytics into practice: The 5th International Workshop on Temporality in Learning Data. *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge* (pp. 488-489). Edinburgh, UK: ACM.

Chen, B., & Zhang, J. (2016). Analytics for knowledge creation: Towards epistemic agency and design-mode thinking. *Journal of Learning Analytics*, *3*(2), 139-163.

Corbett, A. T., & Anderson, J. R. (1994). Knowledge tracing: Modeling the acquisition of procedural knowledge. *User modeling and user-adapted interaction*, *4*(4), 253-278.

Csikszentmihalyi, M., & Larson, R. (2014). Validity and reliability of the experience-sampling method. In *Flow and the foundations of positive psychology* (pp. 35-54). New York: Springer.

Cuban, L. (2001). *Oversold and underused: Computers in the classroom*. Harvard University Press, Cambridge, MA.

Cui, Y., Jin, W. Q., & Wise, A. F. (2017). Humans and machines together: Improving characterization of large scale online discussions through dynamic interrelated post and thread categorization (DIPTiC). *Proceedings of Learning at Scale 2017* (pp. 217-219). Cambridge, MA: ACM.

D'Angelo, C. M., Roschelle, J., & Bratt, H. (2015). Using students' speech to characterize group collaboration quality. *Proceedings of the International Conference on Computer Supported Collaborative Learning*. Gothenburg, Sweden: ISLS.

Dawson, S. (2010). 'Seeing' the learning community: An exploration of the development of a resource for monitoring online student networking. *British Journal of Educational Technology*, *41*(5), 736-752.

Denley, T. (2013). Degree compass: A course recommendation system. *Educause Review Online*.

Dowell, N., Skrypnyk, O., Joksimović, S., Graesser, A. C., Dawson, S., Gašević, D., Vries, P. D., Hennis, T., & Kovanović, V. (2015). Modeling learners' social centrality and performance through language and discourse. *Proceedings of the 8th International Conference on Educational Data Mining* (pp. 250-257). ACM, New York, NY, USA.

Drachsler, H., Verbert, K., Santos, O. C., & Manouselis, N. (2015). Panorama of recommender systems to support learning. In F. Ricci, L. Rokach, B. Shapira and P. B. Kantor (Eds.) *Recommender systems handbook* (pp. 421-451). New York: Springer.

Duval, E., & Verbert, K. (2012). Learning analytics. *E-Learning and Education, 1*(8).

Ertmer, P. A. (1999). Addressing first-and second-order barriers to change: Strategies for technology integration. *Educational Technology Research and Development, 47*(4), 47-61.

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. In R. Vuorikari & J. Castaño Muñoz (Eds.). *Joint Research Centre Science for Policy Report*; EUR 28294 EN; doi:10.2791/955210.

Fritz, J. (2011). Classroom walls that talk: Using online course activity data of successful students to raise self-awareness of underperforming peers. *The Internet and Higher Education*, *14*(2), 89-97.

Gašević, D., Dawson, S., Rogers, T., & Gasevic, D. (2016). Learning analytics should not promote one size fits all: The effects of instructional conditions in predicting academic success. *The Internet and Higher Education*, *28*, 68-84.

Hecking, T., Chounta, I. A., & Hoppe, H. U. (2016). Investigating social and semantic user roles in MOOC discussion forums. *Proceedings of the Sixth International Conference* *on Learning Analytics & Knowledge* (pp. 198-207). ACM New York, NY, USA.

Hero, A. O., & Rajaratnam, B. (2016). Foundational principles for large-scale inference: Illustrations through correlation mining. *Proceedings of the IEEE, 104*(1), 93-110.

Hershkovitz, A., de Baker, R. S. J., Gobert, J., Wixon, M., & Pedro, M. S. (2013). Discovery with models: A case study on carelessness in computer-based science inquiry. *American Behavioral Scientist*, *57*(10), 1480-1499.

Huberth, M., Chen, P., Tritz, J., & McKay, T. A. (2015). Computer-tailored student support in introductory physics. *PloS one*, *10*(9), e0137001.

Jayaprakash, S. M., Moody, E. W., Lauría, E. J., Regan, J. R., & Baron, J. D. (2014). Early alert of academically at-risk students: An open source analytics initiative. *Journal of Learning Analytics, 1*(1), 6-47.

Jeong, H., Biswas, G., Johnson, J., & Howard, L. (2010). Analysis of productive learning behaviors in a structured inquiry cycle using hidden markov models. *Proceedings of the Third International Conference on Educational Data Mining* (pp. 81-90). Pittsburgh, PA: EDM.

Johnson, L., Adams Becker, S., Cummins, M., Estrada, V., Freeman, A., and Hall, C. (2016). *NMC Horizon Report: 2016 Higher Education Edition*. Austin, Texas: The New Media Consortium.

Joksimović, S., Kovanović, V., Jovanović, J., Zouaq, A., Gašević, D., & Hatala, M. (2015). What do cMOOC participants talk about in social media? A topic analysis of discourse in a cMOOC. *Proceedings of the Fifth International Conference on Learning Analytics & Knowledge* (pp. 156-165). ACM.

Joksimović, S., Manataki, A., Gašević, D., Dawson, S., Kovanović, V., & De Kereki, I. F. (2016). Translating network position into performance: Importance of centrality in different network configurations. *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge* (pp. 314-323). ACM.

Klerkx, J., Verbert, K., & Duval, E. (2017). Learning Analytics Dashboards. In C. Lang, G. Siemens, A. Wise & D. Gašević (Eds.) *Handbook of learning analytics* (1st ed, pp. 143-150). Edmonton, AB: SoLAR.

Knight, S., & Littleton, K. (2015). Discourse-centric learning analytics: mapping the terrain. *Journal of Learning Analytics*, *2*(1), 185-209.

Knight, S., Wise, A. F., Chen, B., & Cheng, B. H. (2015). It's about time: 4th international workshop on temporal analyses of learning data. *Proceedings of the Fifth International Conference on Learning Analytics & Knowledge* (pp. 388-389). Poughkeepise, NY: ACM.

Kolb, D. A. (1984). *Experiential education: Experience as the source of learning and learning science*. Englewood Cliffs, NJ: Prentice Hall.

Kovanović, V., Joksimović, S., Gašević, D., Hatala, M., & Siemens, G. (2017) Content analytics: The definition, scope, and an overview of published research. In C. Lang, G. Siemens, A. Wise & D. Gašević (Eds.) *Handbook of learning analytics* (1st ed, pp. 77-92). Edmonton, AB: SoLAR.

Landauer, T. K., MacNamara, D. S., Dennis, S. & Kintsch, W. (Eds.) (2011). *Handbook of latent semantic analysis*. New York, NY: Routledge.

Laney, D. (2001). 3D data management: Controlling data volume, velocity and variety. *META Group Research Note, 6*, 70.

Lockyer, L., Heathcote, E., & Dawson, S. (2013). Informing pedagogical action: Aligning learning analytics with learning design. *American Behavioral Scientist*, *57*(10), 1439-1459.

McNamara, D. S., Crossley, S. A., & McCarthy, P. M. (2010). Linguistic features of writing quality. *Written Communication*, 27(1), 57-86.

McNamara, D., Allen, L., Crossley, S., Dascalu, M., & Perret, C. (2017). Natural Language Processing and Learning Analytics. In C. Lang, G. Siemens, A. Wise & D. Gašević (Eds.) *Handbook of learning analytics* (1st ed, pp. 93-104). Edmonton, AB: SoLAR.

Merceron, A., & Yacef, K. (2008). Interestingness measures for association rules in educational data. *Proceedings for the First International Conference on Educational Data Mining 2008* (pp. 57-66.). Montreal, CA.

Mu, J., Stegmann, K., Mayfield, E., Rosé, C., & Fischer, F. (2012). The ACODEA framework: Developing segmentation and classification schemes for fully automatic analysis of online discussions. *International Journal of Computer-Supported Collaborative Learning*, *7*(2), 285-305.

Nwana, H. S. (1990). Intelligent tutoring systems: An overview. *Artificial Intelligence Review*, *4*(4), 251-277.

Ochoa, X., & Worsley, M. (2016). Augmenting learning analytics with multimodal sensory data. *Journal of Learning Analytics*, *3*(2), 213-219.

Papamitsiou, Z., & Economides, A. (2014). Learning analytics and educational data mining in practice: A systematic literature review of empirical evidence. *Educational Technology & Society, 17*(4), 49–64.

Parameswaran, A., Venetis, P., & Garcia-Molina, H. (2011). Recommendation systems with complex constraints: A course recommendation perspective. *ACM Transactions on Information Systems (TOIS)*, *29*(4), 20.

Pecaric, M., Boutis, K., Beckstead, J., & Pusic, M. (2017). A big data and learning analytics approach to process-level feedback in cognitive simulations. *Academic Medicine*, *92*(2), 175-184.

Poon, L. K., Kong, S. C., Wong, M. Y., & Yau, T. S. (2017). Mining sequential patterns of students' access on learning management system. *International Conference on Data Mining and Big Data* (pp. 191-198). Fukuoka, Japan: Springer.

Poquet, L., & Dawson, S. (2016). Untangling MOOC learner networks. *Proceedings of the Sixth International Conference on Learning Analytics and Knowledge* (pp. 208-212). Edinburgh, UK: ACM.

Prinsloo, P., & Slade, S. (2013). An evaluation of policy frameworks for addressing ethical considerations in learning analytics. *Proceedings of the Third International*

Conference on Learning Analytics and Knowledge (pp. 240-244). Indianapolis, IN: ACM.

Rabbany, R., Takaffoli, M., & Zaïane, O. R. (2011). Analyzing participation of students in online courses using social network analysis techniques. In Pechenizkiy, M., Calders, T., Conati, C., Ventura, S., Romero, C., & Stamper, J. (Eds.), *Proceedings of the 4th International Conference on Educational Data Mining* (pp. 21-30). EDM.

Ritsos, P. D. & Roberts, J. C. (2014). Towards more visual analytics in learning analytics. In M. Phol & J. C. Roberts (Eds.), *Proceedings of the EuroVis Workshop on Visual Analytics* (pp.61-65). Swansea, UK: Eurographics Association.

Roll, I., MacFadyen, L. P., Ni, P., Cimet, M., Shiozaki, L., Paulin, D. & Harris, S. (2016) Questions, not answers: Boosting student participation in MOOC forums. *Proceedings of Learning with MOOC IIIs* (pp 23-25). Philadelphia, PA.

Romero, C., Ventura, S., Pechenizkiy, M., & Baker, R. S. (Eds.). (2010). *Handbook of educational data mining*. New York: CRC press.

Rosé, C. (2017). Discourse Analytics. In C. Lang, G. Siemens, A. Wise & D. Gašević (Eds.) *Handbook of learning analytics* (1st ed, pp. 105-114). Edmonton, AB: SoLAR.

Rosé, C., Wang, Y. C., Cui, Y., Arguello, J., Stegmann, K., Weinberger, A., & Fischer, F. (2008). Analyzing collaborative learning processes automatically: Exploiting the advances of computational linguistics in computer-supported collaborative learning. *International Journal of Computer-Supported Collaborative Learning*, *3*(3), 237-271.

Rummel, N., Walker, E., & Aleven, V. (2016). Different futures of adaptive collaborative learning support. *International Journal of Artificial Intelligence in Education*, *26*(2), 784-795.

Schön, D. A. (1983). *The reflective practitioner: How professionals think in action*. New York, NY: Basic Books.

Sclater, N. (2014). *Code of practice for learning analytics: A literature review of the ethical and legal issues.* Bristol, UK: JISC.

Sclater, N. (2017). *Learning Analytics Explained*. New York: Routledge.

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.

Shneiderman, B. (2014). The big picture for big data: Visualization. *Science*, *343*(6172), 730-730.

Shum, S. B., Knight, S., McNamara, D., Allen, L., Bektik, D., & Crossley, S. (2016). Critical perspectives on writing analytics. *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge* (pp. 481-483). Edinburgh, UK: ACM.

Siemens, G., Gašević, D., Haythornthwaite, C., Dawson, S., Buckingham Shum, S., Ferguson, R., Duval, E., Verbert, K., & Baker, R. S. (2011). *Open learning analytics: An integrated & modularized platform*. [Concept paper]. Society for Learning Analytics Research.

Sinha, T., Jermann, P., Li, N & Dillenbourg, P. (2014). Your click decides your fate: Inferring information processing and attrition behavior from MOOC video clickstream interactions. *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP) Workshop on Modeling Large Scale Social Interaction in Massively Open Online Courses*, (pp. 3-14). Doha, Qatar: ACL.

Slade, S. & Prinsloo, P. (2014). Student perspectives on the use of their data: Between intrusion, surveillance, and care. In *Challenges for research into open & distance learning: Doing things better – doing better things* (pp. 291–300). Oxford, UK: European Distance and E-Learning Network.

Slade, S., & Prinsloo, P. (2013). Learning analytics: Ethical issues and dilemmas. *American Behavioral Scientist*, *57*(10), 1510-1529.

Suthers, D. D., & Desiato, C. (2012). Exposing chat features through analysis of uptake between contributions. *Proceedings of the 45th Hawaii International Conference on System Science* (pp. 3368–3377). Maui, HI: IEEE.

Suthers, D., Wise, A. F., Schneider, B., Shaffer, D. W., Hoppe, H. U. & Siemens, G. (2015). Learning analytics of and in mediational processes of collaborative learning. In O. Lindwall, P. Häkkinen, T. Koschmann, P. Tchounikine & S. Ludvigsen (Eds.), *Proceedings of the Eleventh International Conference on Computer Supported Collaborative Learning* (Vol I, pp. 26-30). Gothenburg, Sweden: ISLS.

Svihla, V., Wester, M. J., & Linn, M. C. (2015). Distributed revisiting: An analytic for retention of coherent science learning. *Journal of Learning Analytics*, *2*(2), 75-101.

van Leeuwen, A. (2015). Learning analytics to support teachers during synchronous CSCL: Balancing between overview and overload. *Journal of Learning Analytics*, *2*(2), 138-162.

Velazquez, E., Ratté, S., & de Jong, F. (2016). Analyzing students' knowledge building skills by comparing their written production to syllabus. *Proceedings of the International Conference on Interactive Collaborative Learning* (pp. 345-352). Belfast, UK: Springer.

Verbert, K., Duval, E., Klerkx, J., Govaerts, S., & Santos, J. L. (2013). Learning analytics dashboard applications. *American Behavioral Scientist*, *57*(10), 1500-1509.

Vytasek, J., Wise, A. F., & Woloshen, S. (2017). Topic models to support instructors in MOOC forums. *Proceedings of the Seventh International Conference on Learning Analytics & Knowledge* (pp. 610-611). Vancouver, CA: ACM.

Whitelock, D., Twiner, A., Richardson, J. T., Field, D., & Pulman, S. (2015,). OpenEssayist: A supply and demand learning analytics tool for drafting academic essays. *Proceedings of the Fifth International Conference on Learning Analytics & Knowledge* (pp. 208-212). Poughkeepsie, NY: ACM.

Winne, P. H. (2010). Improving measurements of self-regulated learning. *Educational Psychologist*, *45*(4), 267-276.

Winne, P. H. (in press). Leveraging big data to help each learner upgrade learning and accelerate learning science. *Teachers College Record.*

Wise, A. F. & Chiu, M. M. (2011). Analyzing temporal patterns of knowledge construction in a role-based online discussion. *International Journal of Computer-Supported Collaborative Learning*. *6*(3), 445-470.

Wise, A. F. & Shaffer, D. W. (2015). Why theory matters more than ever in the age of big data. *Journal of Learning Analytics, 2*(2), 5-13.

Wise, A. F. & Vytasek, J. M. (2017). Learning analytics implementation design. . In C. Lang, G. Siemens, A. Wise & D. Gašević (Eds.) *Handbook of learning analytics* (1st ed, pp. 151-160). Edmonton, AB: SoLAR.

Wise, A. F. Vytasek, J. M., Hausknecht, S. N. & Zhao, Y. (2016). Developing learning analytics design knowledge in the "middle space": The student tuning model and align design framework for learning analytics use. *Online Learning, 20(2),* 1-28.

Wise, A. F., Cui, Y. & Jin, W. Q. (2017). Honing in on social learning networks in MOOC forums: Examining critical network definition decisions. *Proceedings of the Seventh International Conference on Learning Analytics & Knowledge* (pp. 383-392). Vancouver, CA: ACM.

Wise, A. F., Speer, J., Marbouti, F. & Hsiao, Y. (2013). Broadening the notion of participation in online discussions: Examining patterns in learners' online listening behaviors. *Instructional Science. 41*(2), 323-343.

Wise, A. F., Zhao, Y. & Hausknecht, S. N. (2014). Learning analytics for online discussions: Embedded and extracted approaches. *Journal of Learning Analytics, 1*(2), 48-71*.*

Yang, D., Sinha, T., Adamson, D., & Rose, C. P. (2013). Turn on, tune in, drop out: Anticipating student dropouts in massive open online courses. *Proceedings of the 2013 NIPS Workshop on Data-Driven Education*. NIPS Foundation.

Zheng, A. (2015). *Evaluating machine learning models*. Boston, MA: O'Reilly Media.

Zhu, M., Bergner, Y., Zhang, Y., Baker, R. S. J. D., Wang, Y., Paquette, L., & Barnes, T. (2016). Longitudinal engagement, performance, and social connectivity: A MOOC case study using exponential random graph models. *Proceedings of the Sixth International Conference on Learning Analytics and Knowledge* (pp. 223-230). Edinburgh, UK: ACM.