Piecing the Learning Analytics Puzzle: A Consolidated Model of a Field of Research and Practice

Dragan Gašević[1,2,*], Vitomir Kovanović[2], and Srećko Joksimović[1]

[1]Moray House School of Education
[2]School of Informatics
The University of Edinburgh, United Kingdom
{dragan.gasevic, v.kovanovic, s.joksimovic}@ed.ac.uk

Abstract. The field of learning analytics was founded with the goal to harness vast amounts of data about learning collected by the extensive use of technology. After the early formation, the field has now entered the next phase of maturation with a growing community who has an evident impact on research, practice, policy, and decision-making. Although learning analytics is a bricolage field borrowing from many related other disciplines, there is still no systematized model that shows how these different disciplines are pieced together. Existing models and frameworks of learning analytics are valuable in identifying elements and processes of learning analytics, but they insufficiently elaborate on the links with foundational disciplines. With this in mind, this paper proposes a consolidated model of the field of research and practice that is composed of three mutually connected dimensions – theory, design, and data science. The paper defines why and how each of the three dimensions along with their mutual relations is critical for research and practice of learning analytics. Finally, the paper stresses the importance of multi-perspective approaches to learning analytics based on its three core dimensions for a healthy development of the field and a sustainable impact on research and practice.

Keywords. Learning analytics, educational research, learning design, data science, and interaction design

1 Introduction

Learning analytics is a field developed to harness unprecedented amounts of data collected by the extensive use of technology in education. The formation of learning analytics brought together researchers and practitioners from a wide range of fields such as education, psychology, economics, statistics, data mining, and information visualization. Although individuals from all these domains had previously worked with data in what is now commonly referred to as learning analytics, they did not have a joint forum for exchanging ideas, methods, and results. The announcement of the first International Conference on Learning Analytics (LAK) held in Canada in 2011 can probably be marked as the official point for the field formation (Long, Siemens, Conole, & Gašević, 2011). Since then the field has come a long way which is well demonstrated through a number of achievements such . For example, the proceedings of the LAK conference are among the top 8 most cited publication in education

[*] Corresponding author: School of Informatics, The University of Edinburgh, 10 Crichton St, Edinburgh EH8 9AB, United Kingdom. Phone: +44 131 651 3837

technology (Google Scholar, 2016). The Society for Learning Analytics Research (SoLAR)[1], as a leading professional organization in learning analytics, have developed a strong influence among a wide range of organization and stakeholders. The field has a specialized journal with a rapidly rising reputation – Journal of Learning Analytics[2]. There is an event growth in the number of doctoral students and quality of their work (Pechenizkiy & Gašević, 2014); Finally, many education institutions, systems, and funding bodies have made significant investments into learning analytics; and a board network of events and organizations in all parts of the world[3].

Learning analytics is an *interdisciplinary* field of *practice* and *research* that takes a *holistic* approach to employing data to address questions of relevance for learning, teaching, and education(Siemens & Gasevic, 2012). The commonly used definition, adopted by SoLAR, defines learning analytics as "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" (Long et al., 2011). The purpose of analytics – both for understanding and optimization of learning – implies practical and research components of the field. The holistic nature is recognized through the study of learning context and environments. Finally, the interdisciplinary nature is demonstrated through a broad representation of different fields in learning analytics publications, events, and initiatives (Dawson, Gašević, Siemens, & Joksimovic, 2014).

Although it is widely accepted that learning analytics is a bricolage field borrowing from many related disciplines, there is still no consolidated model that systematizes how these different disciplines are pieced together. The paper aims to address this gap in the literature and proposes a consolidated model of learning analytics that recognizes its key characteristics – field of research and practice, holistic in nature, and interdisciplinarity. The model synthesizes existing results in learning analytics and suggests that learning analytics is composed of three mutually connected dimensions – theory, design, and data science. To illustrate why the consideration of the three dimensions is critical for learning analytics, the paper starts with a critical interrogation of the scope, topics, and challenges of learning analytic. The paper then goes on and reviews the dimensions of the model and their mutual relationships, and finally, concludes with a discussion of the implications for future research and practice.

2 Learning Analytics: Scope, Topics, and Challenges

Learning analytics encompasses a wide range of activities that are broadly associated with the four core elements of learning analytics – collection, measurement, analysis, and reporting – identified in the definition of learning analytics (Long et al., 2011). Reimann (2016) notes that, while the use of data in educational research and learning sciences has been present for quite some time, learning analytics differs from the "traditional" data analyses in education as it focuses on the longitudinal collection of a large number of data points from authentic learning environments. Although the field of learning analytics supports the collection of a wide range of data, a bulk of existing work in learning analytics is dedicated to digital traces collected through the interaction of people with technology, content, and/or

[1] http://solaresearch.org
[2] http://learning-analytics.info
[3] http://lasi.solaresearch.org

other people (Siemens & Gasevic, 2012). An important task of learning analytics is the development of measures that can a) offer practical insights into learning processes and outcomes, and b) be theoretically interpreted (Gašević, Dawson, & Siemens, 2015). For analysis, learning analytics adopts a wide range of methods from the fields such as data mining, statistics, social network analysis, process mining, and text analysis. Exchange of analysis methods is also a point of active collaboration between educational data mining and learning analytics (Baker & Siemens, 2014; Siemens & Baker, 2012). Finally, reporting in learning analytics aims to provide actionable insight to a broad range of stakeholders by building on principles of information visualization and human-computer interaction (Verbert, Duval, Klerkx, Govaerts, & Santos, 2013).

Learning analytics covers a number of different topics of research and practical significance for learning, teaching, and education (Dawson et al., 2014; Ferguson, 2012). Without any pretension to provide a systematic and comprehensive overview of the literature, we only highlight some of the themes that attracted much attention in the current literature. Prediction of student retention and learning outcomes are among the most popular topics in learning analytics due to their practical relevance for educational institutions (Jayaprakash, Moody, Lauría, Regan, & Baron, 2014). Analysis of 21st-century skills (Buckingham Shum & Deakin Crick, 2016), self-regulated learning (Roll & Winne, 2015), and social learning (Dowell et al., 2015) are topics gaining much attention in learning analytics. Research on topics such as learning behavior (Käser, Klingler, & Gross, 2016), strategies (Kovanović, Gašević, Joksimović, Hatala, & Adesope, 2015), dispositions (Buckingham Shum & Deakin Crick, 2012), and affective states (D'Mello, 2017) are commonly used to inform designs of learning analytics presentation to stakeholders (Verbert et al., 2013). Privacy protection (Drachsler & Greller, 2016), ethics (Ferguson, Hoel, Scheffel, & Drachsler, 2016), policy (Macfadyen & Dawson, 2012), infrastructure (Apereo Foundation, 2016), and standards (Bakharia, Kitto, Pardo, Gašević, & Dawson, 2016) are also key themes associated with the systemic adoption of learning analytics in educational institutions (Ferguson et al., 2014).

Existing work in learning analytics provides some proposals for models of learning analytics that are either identify dimensions of learning analytics or process phases or both of the previous two. A reference model of learning analytics is suggested by Chatti, Dyckhoff, Schroeder, & Thüs (2012) that identifies four key dimensions of learning analytics each one dedicated to the following four questions: what – data, environment, and context; why – objectives; how – methods; and who – stakeholders. Similarly, these four dimensions, with somewhat different names and scope, are also recognized in a generic framework for learning analytics proposed by Greller & Drachsler (2012). The Greller & Drachsler generic framework also includes internal limitations (i.e., competencies and acceptance) and external constraints (i.e., norms and conventions). An interactive process model of learning analytics is suggested by Steiner, Kickmeier-Rust, & Türker (2014) through a synthesis of several other process models and consist of the following phases: data selection, data capturing, data aggregation, data reporting, prediction, acting upon results, and refinement. Finally, Scheffel, Drachsler, Stoyanov, and Specht (2014) propose a framework that recommends quality indicators for learning analytics organized across dimensions commonly identified in the existing models/frameworks.

Although much growing excitement around learning analytics is evident, there are also several challenges that the field is grappling to address. For example, concerns are often reported in connection of the relevance of the ways how predictive models are constructed to inform teaching practice (Gašević, Dawson, Rogers, & Gasevic, 2016). The accuracy of predictive models, when applied in different contexts, is found to decline (Jayaprakash et al., 2014) or to identify predictors that are not relevant for teaching practice (Gašević et al., 2016). Evaluation of the effectiveness of learning analytics dashboards with learners in practice is rare (Verbert et al., 2013) and even when available, results of some of the existing studies are not positive. For example, adverse effects of learning analytics dashboards on the intrinsic motivation of undergraduate students were also reported (Krumm, Waddington, Teasley, & Lonn, 2014). Difficulties in interpretation of learning analytics dashboards commonly available in learning management systems by learners are documented (Corrin & de Barba, 2014). Some early warning alert systems are shown to promote suboptimal teaching practice (Tanes, Arnold, King, & Remnet, 2011), even when they generate profits in student retention (Arnold & Pistilli, 2012). The methodological validity of some of the studies on the use of learning analytics has been challenged in public debates (Caulfield, 2013) and such challenges are supported by recent results of replications studies (Dawson, Jovanović, Gašević, & Pardo, 2017).

We posit that much of the above issues can be attributed to the shortage of guidelines that are grounded in foundational disciplines of learning analytics and that should inform activities related to different dimensions of research and practice in learning analytics. Existing models and frameworks, discussed above, are valuable contributions that go beyond the definition of learning analytics and offer possible operationalizations of what learning analytics research and practice should involve. However, they insufficiently detail links between the principles established in disciplines on which learning analytics builds as a bricolage field such as educational research, learning science, psychology, human-computer interaction, data mining, or research methods. This paper aims to address this gap in the existing literature by proposing a consolidated model of learning analytics.

3 Consolidated Model of Learning Analytics

We posit that the foundational principles of learning analytics can be grouped around three mutually connected dimensions – theory, design, and data science. We also posit that the most effective results in and with highest validity for research and practice can be achieved only once the principles of all three dimensions are considered. The consolidated model (Figure 1) does not exclude the existing models and frameworks of learning analytics, but rather complements them.

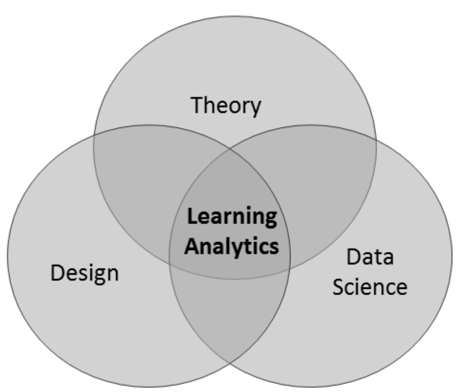

Figure 1. Consolidated model of learning analytics

3.1 Theory

As a field that seeks to contribute to broad educational research and practice, the theory in learning analytics needs to play a pivotal role. Early adoption of data-driven approaches in research and practice of learning, teaching, and education sparked much enthusiasm and shed some light on the power of data science methods (Romero & Ventura, 2010). However, data-driven applications were soon recognized as not sufficiently informative for research and practice of learning analytics (Gašević et al., 2015). Such weaknesses in data-driven approaches led to the emphasis of theory-orientation of learning analytics as one of the fundamental differences with other mainstream data-driven approaches (Baker & Siemens, 2014). In his critique of preliminary advocates of data-driven analytics (e.g., Andreson (2008)) , Reimann (2016) suggests that such atheoretical approaches to learning analytics are a misconception what scientific method is and “the logical (and ethical) error of using descriptions of the past as prescriptions for the future” (ibid, p. 136).

There is a growing agreement on theory-orientation as critical for research and practice of learning analytics. Theory orientation has been recognized as essential for informing the choice of questions asked and hypotheses tested (Rogers, Gašević, & Dawson, 2016; Wise & Shaffer, 2015). Theory is suggested as a driving aspect for the selection of methods used for study design and data analysis (Reimann, 2016; Rogers et al., 2016). Various scholars also argue that results from the learning analytics research should be interpreted with respect to and inform existing theory (Gašević et al., 2015; Reimann, 2016; Wise & Shaffer, 2015). Finally, producing actionable insights for practitioners (Gašević et al., 2016; Rogers et al., 2016) and designing environments and tools that integrate learning analytics for different stakeholders (Marbouti & Wise, 2016) should be theoretically grounded.

Potential to analyze and identify patterns in large data sets is often associated with learning analytics. While highly promising to include large population numbers with unprecedented numbers of data points, this also results in a high statistical power where many associations, even those with very small effect sizes (Cohen, 1992), can be detected as significant (Wise & Shaffer, 2015). The role of theory in such cases is to identify which of these associations are meaningful and include them into analytical models as hypotheses to be tested. Theory also allows for validating if associations between some learning-related constructs extracted from digital traces and learning outcomes should be even considered valid. In some cases, there is no theoretically justified relation between a digital trace, which represents a record of an event about the use of a learning tool, and a learning outcome. Rather, the tool use is provided in the environment to trigger a learning process, which can then facilitate reaching particular outcomes (Reimann, 2016). In such cases, if data about learning processes is not collected, we might not be able to understand if certain learning processes are activated, and if so, under what conditions and what learning outcomes are associated with different conditions. Likewise, data about context is rarely used in learning analytics, even though context critically shapes learning according to contemporary theories (Winne, 2006). Such imperfect practices have direct implications for the current endeavors in learning analytics to collect data about activation of learning processes; the collection of data about learning processes is frequently not the case. Data about learning processes can be gathered through the use of established self-report instruments (Lust, Elen, & Clarebout, 2013), theory-based instrumentation of learning environments (Siadaty, Gašević, & Hatala, 2016; Zhou & Winne, 2012), discourse (Kovanović et al., 2016), psychophysiological measures (Azevedo, 2015), and different qualitative methods.

Adoption of theory can help explain the reasons for inconsistencies in the results across several studies and clarify psychological, sociological, contextual, cultural, instructional, and other mechanisms that affected the results. For example, building on Winne and Hadwin's (1998) model of self-regulated learning, Gašević and colleagues (2016, 2015; 2017) showed that accounting for external (e.g., instructional design) and internal (e.g., motivation and prior knowledge) conditions is essential for interpretation of the results of analytics in order to inform both research and practice. Failing to accounts for contextual conditions (in a broad sense) could lead towards misinterpretation of findings and limit replication of existing research in different settings (Gašević et al., 2016). Similarly, Joksimović and colleagues (2016) identified a problem of inconsistent results in the association of learning performance with centrality measures in social networks, even though this association can theoretically be justified. Joksimović et al. theorized and empirically showed that only if the ties in networks are weak, the relationship can be established, as also theoretically justified (Granovetter, 1982).

Sources of theory in learning analytics are not limited only to psychology, but they also include disciplines such as sociology, organization science, information science, and semiotics. Although both design and data science have their own theoretical foundations, theories that are essential for learning analytics are about human learning, teaching, and education. Connections of theory with design and data science methods are further discussed in the remaining two sections.

3.2 Design

As a field of research and practice, learning analytics is intrinsically linked with different forms of design. Design in analytics concerns three main dimensions – interaction and visualization design, learning design, and study design.

3.2.1 Interaction and visualization design

Much work on learning analytics is related to the development of interactive visualizations that can inform the decision-making of stakeholders. The literature often refers to the designs of different interactive visualizations with the intent to support self-regulation and awareness (Verbert et al., 2013). However, the learning analytics literature offers limited reasoning on how specifically an interactive visualization will support self-regulation by building on theories and research of self-regulated learning. There is also little empirical evidence in learning analytics that proves a positive impact of such visualizations on self-regulation and more specifically that pins down to any specific process of self-regulated learning that is triggered by such visualizations.

Much of the existing limitations of the design of interactive visualizations are due to the widespread assumption that just an act of visually presenting something will promote desirable learning and that the use of visualization will ease the interpretation of analytical results. We suggest that the decision of learning analytics needs to build on the existing principles and theories established in technology acceptance, interaction design, and learning science. The theory of cognitive fit does not give benefits of one information presentation over another one in general, but rather suggests that the presentation should be the one that offers the best cognitive fit for a given task (Vessey, 1991). The theory of technology-task fit (Goodhue & Thompson, 1995) suggests that the usefulness of a technology should not be sacrificed solely for the purpose of making the technology easier to use. Users can tolerate a lesser ease of use, if the technology is highly useful, which applies in particular to the technology that supports complex task as commonly the case in learning. For example, while ease to understand, the metaphor of traffic lights in Course Signals provided insufficient information to teaching staff to promote the provision of feedback based on established principles in educational literature (Gašević et al., 2015; Tanes et al., 2011).

We posit that interaction design in general and design of interactive visualizations in particular in learning analytics needs to be a) grounded in learning theory and b) tailored specifically to promote and evaluate the activation of a particular a (set of) learning mechanism(s). Just building on the established principles and mantras for information visualizations and interaction design is criticized as insufficient to create a body of research knowledge and guide evaluations even in the field of information visualization (Liu, Nersessian, & Stasko, 2008). Wise and her colleagues (2016; 2014) give a good example how the design of interactive visualization can be effectively conducted in learning analytics. Wise et al. build on the existing literature in computer supported collaborative learning that identifies a bias of the most recent discussion posts as being only considered by learners, instead of reading entire threads. This bias constraints the level and quality of knowledge construction, as students do not take advantage of the richness and diversity of complete discussions. Wise and her colleagues proposed two approaches to interactive visualizations (embedded and extracted) (Wise et al., 2014). The empirical validation of this method looked at specific knowledge construction processes activated as a consequence of the use of the interactive visualizations (Marbouti & Wise, 2016).

3.2.2 Learning design

Learning analytics need to a) account for the effects of learning design in order to achieve theoretically valid results, b) select relevant data and appropriate analysis methods, and c) give actionable insights that can inform practice. As already discussed, if the moderating effect of learning design is not considered, predictive models of academic performance cannot reveal factors of significance for teaching practice (Gašević et al., 2016). In such cases, the choices of variables that are used should be driven by the decisions made in learning designs (Lockyer, Heathcote, & Dawson, 2013), including tasks given to learners, tools offered, and collaborations with other people envisioned (Goodyear, 2015). An aggregated analysis of data collated across different courses needs to account for the effects of variance across various groups, where variance is induced by differences in learning designs. In such cases, the nested nature of data should be reflected in the choice of methods that are used for analysis.

Learning analytics need to be integrated into learning designs. Rather than introducing learning analytics just because they are available in the learning environment, a learning design should carefully consider the purpose of the inclusion of learning analytics and their role in connection to specific objectives and tasks in the learning design (Lockyer et al., 2013). Wise (2014) offers a theoretically-grounded and practice-oriented integration of learning analytics with learning design that can serve as an excellent source for the future work. Wise proposes a model that aims to a) promote the development of learners’ agency through personal goal setting and reflection empowered by analytics, b) offer scaffolding for the agency development through reference frames delivered by learning analytics, and c) foster a dialogue between learners and teaching staff, similar to the notion of visible learning proposed by Hattie (2012), where learning analytics serves as a conversation opener.

3.2.3 Study design

Although high numbers of data points are collected longitudinally in authentic learning settings, learning analytics cannot be removed from the careful consideration of study design. As in conventional educational research, study design is tightly associated with theoretical framing, data collection, analysis, methods, and interpretation. Omissions in the interpretation of the results such as causal vs. correlational can easily be made without consideration of the study design. A study design needs to be outlined even when historical data, routinely collected through operation of learning technology, are analyzed. A study design consideration should understand the nature of data collection, possible ways how the data can be analyzed, and the types of questions to be answered. Evaluations that test effects of learning analytics-based interventions need to have a strong emphasis on study design in order to avoid partial interpretations (Caulfield, 2013) and promote replication (Dawson et al., 2017).

Design-based research is suggested to have a 'natural' fit with learning analytics, as both are used to inform practice and advance theory (Reimann, 2016). Design-based research is an iterative research method in which each iteration aims to address practical issues that from the previous iteration through the use of technological and pedagogical interventions, which are in turn grounded in theory (Anderson & Shattuck, 2012). Therefore, each intervention can be seen as a validation of a theory, while the practical solutions are improved iteratively by typically following quasi-experimental study designs. Context in design-based research (and hence, learning analytics) is considered an intrinsic part of interventions. Therefore, contextual data need to be accounted for in learning analytics "as [they are] providing the resources through which theoretically expected learning processes become realized in a specific learning situation" (Reimann, 2016, p. 137).

3.3 Data Science

Methods and techniques of data science are an essential part of learning analytics. They enable the main four phases mentioned in the definition of learning analytics including collection, measurement, analysis, and reporting. A comprehensive classification of data science methods based on the types of questions they can address is provided by Steiner et al. (2014) and includes: prediction models – e.g., classification, regression, and latent knowledge estimation; structure discovery – clustering, factor analysis, outlier detection, domain structure discovery, and social network analysis; relationship mining – association rule mining, correlational mining, sequential pattern mining, causal data mining; and other approaches – process mining, discourse analysis, and multimodal approaches. A comprehensive overview of the tools commonly used in educational data mining, a sister field from which learning analytics draws methods, is provided by Slater, Joksimović, Kovanović, Baker, and Gašević (in press).

Although the use of data science methods is assumed in learning analytics, some issues in their application need to be highlighted especially in connection to the two other dimensions of the model – theory and design. As already established, learning analytics is theory-oriented, and thus, the models that are used for analysis need to support theoretical foundations used in different cases. Theory-orientation typically refers to the choice of methods and the interpretation of the results. In the previously mentioned study, Joksimović et al. (2016) showed that standard methods of descriptive social network analysis (SNA) (Wasserman, 1994) are not sufficient to test certain network characteristics such as the propensity of strong ties, reciprocity, and homophilic relationships, which require more advanced statistical SNA methods such as exponential random graph models (Lusher, Koskinen, & Robins, 2012). The example of nested nature of data as a consequence of learning design has already been mentioned in Section 3.2.2.

Data science methods should recognize that the notion of complex systems is gaining much attention in theoretical underpinnings of learning (Jacobson, Kapur, and Riemman, 2016) and education (Macfadyen et al., 2014). As noted by Reimann (2016), general regression models are insufficient to address the structure and nature of complex systems. For example, referring back to Wise's (2014) proposal for theory grounded integration of learning design and learning analytics, the role of learning analytics in both analysis methods and study design need to be recognized. This is particularly significant given that learning analytics, embedded into learning designs, seek to establish feedback loops between students and teaching staff that did not exist otherwise. Therefore, the application of approaches such as system dynamics as used for understanding of complex non-linear systems is an important direction that has been underexplored in learning analytics.

Data science needs to provide methods that account for multiple dimensions of relevance for the study of learning rather than just using counts of clickstreams. Gašević and colleagues (2015) suggest that Winne's (1997) COPES model (conditions, operations, productions, evaluation, and standards) can serve as an excellent theoretical foundation for the adoption, refinement, and development of data science methods in learning analytics. For example, to analyze psychophysiological measures of learning, a combination of digital signal processing and machine learning methods is necessary to obtain insights that can inform theory and practice (Pijeira-Díaz, Drachsler, Järvelä, & Kirschner, 2016). Text analysis methods are particularly useful to analyze discourse in connection to cognitive, metacognitive, affective, and motivational aspects of learning (Azevedo, 2015), and thus, for the study of conditions, operations, and products of the COPES model. Finally, to examine learning as a process of temporal and developmental nature, methods developed in the process and sequence mining areas can be particularly useful (Molenaar, 2014).

4 Discussion and Future Consideration

The main contribution of the consolidated model proposed in this paper lies in the guidance about key dimensions that need to be taken into consideration in research and practice of learning analytics. Consideration of only one or even two dimensions of the consolidated model is insufficient to unlock the full potential and assure the validity of learning analytics research and practice. The inclusion of only theory and design is not part of learning analytics as it does not make use of any data. The intersection of theory and data science is relevant, but the lack of design deliberation cannot provide sufficient practical or research validity. Finally, the intersection of design and data science is also not enough without the well-established theoretical foundations of learning and teaching.

Although all three dimensions – theory, design, and data science – should be part of learning analytics research and practice, not all of their sub-dimensions are necessary in every situation. For instance, we do not have to always include all sub-dimensions of design. If the focus of learning analytics research and practice is on the *understanding* part of the learning analytics definition (Long et al., 2011), it is not necessary to ponder on interaction and/or visualization design; unless, the understanding part seeks to understand effects of a visualization on learning, teaching, or education. In other cases, if a research study, which collates digital traces and data science methods for analysis, is conducted in a controlled setting in which students did not use a specific learning design and/or visualization, it is not necessary to reflect on learning design and/or interaction design. A special case is design-based research in which learning design, as pedagogical interventions and context, is always an essential component of study design. However, the study design is a required sub-dimension of learning analytics research as well as any practical application that aims to have a robust evaluation.

While in actual implementations the theory dimension is less dominant, it is still recommended to ground practice in strong (theoretical) principles of learning, teaching, and education. This relates to data science and all three sub-dimensions of the design. Learning designs should be based on accepted learning theories and constructs such as cognitive load (Sweller, 1994) or desirable difficulties (Bjork & Bjork, 2011). Interaction design should be more than just attractive visualizations and incorporate elements of theory pertinent to the purpose a visualization aims to achieve such as support for self-regulated learning or computer supported collaborative learning (Marbouti & Wise, 2016; Wise, 2014). The study design should always be guided by the theoretical considerations as otherwise, the validity of findings for both research and practice could be severely compromised.

The use of different data science methods and systems is an essential dimension of any research and practice in learning analytics. The choice of methods and systems is driven by theory and design in both research and practice of learning analytics. Results produced by the use of data science methods and systems should, in turn, inform theory and design.

The consolidated model of learning analytics, proposed in this paper complements existing models and frameworks of learning analytics (Chatti et al., 2012; Greller & Drachsler, 2012; Steiner et al., 2014) as well as other critical considerations in learning analytics such as privacy, ethics, policy, literacy, and standards (Bakharia et al., 2016; Ferguson et al., 2014; Tsai & Gašević, in press; Wasson & Hansen, 2016). The model recommends how theory, design, and data science a) mutually interact and b) can inform dimensions of the existing models. For example, the what, why, how, and who questions of the Chatti et al. (2012), which are in some forms available in the Greller & Drachsler model (2012), need to be informed by theory, design, and data science. Finally, we posit that the three dimensions – theory, design, and data science – are enabled by and can further inform considerations related to privacy, ethics, policy, literacy, and standards.

The consolidated model of learning analytics suggested in this paper should inform the further research and practice of learning analytics. We expect that the dimensions provided in the model can assist researchers in their deliberations what elements their studies should include, from the early phases of conceptualization to the final stages of empirical validation and interpretation of the results. The model is also relevant to practitioners in informing their decisions regarding the choices of learning analytics tools, their integration in existing practice, and evaluation of learning analytics applications. The proposed model can assist in the development of the field of learning analytics by offering foundational dimensions around which existing and future contributions from research and practice will be systematized and critically interrogated.

References

Anderson, T., & Shattuck, J. (2012). Design-Based Research A Decade of Progress in Education Research? *Educational Researcher*, *41*(1), 16–25. https://doi.org/10.3102/0013189X11428813

Andreson, C. (2008). The End of Theory: The Data Deluge Makes the Scientific Method Obsolete. Retrieved January 13, 2017, from https://www.wired.com/2008/06/pb-theory/

Apereo Foundation. (2016). Learning Analytics Initiative | Apereo. Retrieved January 5, 2017, from https://www.apereo.org/communities/learning-analytics-initiative

Arnold, K. E., & Pistilli, M. D. (2012). Course signals at Purdue: using learning analytics to increase student success. In *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge* (pp. 267–270). New York, NY, USA: ACM. https://doi.org/10.1145/2330601.2330666

Azevedo, R. (2015). Defining and Measuring Engagement and Learning in Science: Conceptual, Theoretical, Methodological, and Analytical Issues. *Educational Psychologist*, *50*(1), 84–94. https://doi.org/10.1080/00461520.2015.1004069

Baker, R., & Siemens, G. (2014). Educational data mining and learning analytics. In R. K. Sawyer (Ed.), *Cambridge Handbook of the Learning Sciences:* Cambridge, UK: Cambridge University Press.

Bakharia, A., Kitto, K., Pardo, A., Gašević, D., & Dawson, S. (2016). Recipe for Success: Lessons Learnt from Using xAPI Within the Connected Learning Analytics Toolkit. In *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge* (pp. 378–382). New York, NY, USA: ACM. https://doi.org/10.1145/2883851.2883882

Bjork, E. L., & Bjork, R. A. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. In M. A. Gernsbacher, R. W. Pew, L. M. Hough, & J. R. Pomerantz (Eds.), *Psychology and the real world: Essays illustrating fundamental contributions to society* (pp. 56–64).

Buckingham Shum, S., & Deakin Crick, R. (2012). Learning Dispositions and Transferable Competencies: Pedagogy, Modelling and Learning Analytics. In *Proceedings of the 2Nd International Conference on Learning Analytics and Knowledge* (pp. 92–101). New York, NY, USA: ACM. https://doi.org/10.1145/2330601.2330629

Buckingham Shum, S., & Deakin Crick, R. (2016). Learning Analytics for 21st Century Competencies. *Journal of Learning Analytics*, *3*(2), 6–21. https://doi.org/10.18608/jla.2016.32.2

Caulfield, M. (2013, November 12). What the Course Signals “Kerfuffle” is About, and What it Means to You. Retrieved from http://www.educause.edu/blogs/mcaulfield/what-course-signals-kerfuffle-about-and-what-it-means-you

Chatti, M. A., Dyckhoff, A. L., Schroeder, U., & Thüs, H. (2012). A reference model for learning analytics. *International Journal of Technology Enhanced Learning*, *4*(5–6), 318–331.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155–159.

Corrin, L., & de Barba, P. (2014). Exploring students’ interpretation of feedback delivered through learning analytics dashboards. In *Proceedings of the ascilite 2014 conference* (pp. 629–633). Dunedin, NZ.

Dawson, S., Gašević, D., Siemens, G., & Joksimovic, S. (2014). Current State and Future Trends: A Citation Network Analysis of the Learning Analytics Field. In *Proceedings of the Fourth International*

Conference on Learning Analytics And Knowledge (pp. 231–240). New York, NY, USA: ACM. https://doi.org/10.1145/2567574.2567585

Dawson, S., Jovanović, J., Gašević, D., & Pardo, A. (2017). From prediction to impact: Evaluation of a learning analytics retention program. In *Proceedings of the 7th International Conference on Learning Analytics and Knowledge (LAK 2017)*. Vancouver, BC, Canada: ACM Press.

D'Mello, S. K. (2017). Emotional Learning Analytics. In C. Lang, G. Siemens, W. Alyssa, & D. Gašević (Eds.), *Handbook of Learning Analytics & Educational Data Mining* (First edition). Edmonton, AB, Canada: Societz for Learning Analytics Research.

Dowell, N. M., Skrypnyk, O., Joksimovic, S., Graesser, A. C., Dawson, S., Gaševic, D., … Kovanovic, V. (2015). Modeling Learners' Social Centrality and Performance through Language and Discourse. In *Proceedings of the 8th International Educational Data Mining Society* (pp. 250–257). Madrid, Spain: IEDMS.

Drachsler, H., & Greller, W. (2016). Privacy and Analytics: It's a DELICATE Issue a Checklist for Trusted Learning Analytics. In *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge* (pp. 89–98). New York, NY, USA: ACM. https://doi.org/10.1145/2883851.2883893

Ferguson, R. (2012). Learning analytics: drivers, developments and challenges. *International Journal of Technology Enhanced Learning*, *4*(5), 304–317. https://doi.org/10.1504/IJTEL.2012.051816

Ferguson, R., Hoel, T., Scheffel, M., & Drachsler, H. (2016). Guest Editorial: Ethics and Privacy in Learning Analytics. *Journal of Learning Analytics*, *3*(1), 5–15. https://doi.org/10.18608/jla.2016.31.2

Ferguson, R., Macfadyen, L. P., Clow, D., Tynan, B., Alexander, S., & Dawson, S. (2014). Setting Learning Analytics in Context: Overcoming the Barriers to Large-Scale Adoption. *Journal of Learning Analytics*, *1*(3), 120–144.

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 learning success. *The Internet and Higher Education*, *28*, 68–84. https://doi.org/10.1016/j.iheduc.2015.10.002

Gašević, D., Dawson, S., & Siemens, G. (2015). Let's not forget: Learning analytics are about learning. *TechTrends*, *59*(1), 64–71. https://doi.org/10.1007/s11528-014-0822-x

Gašević, D., Mirriahi, N., Dawson, S., & Joksimović, S. (2017). Effects of instructional conditions and experience on the adoption of a learning tool. *Computers in Human Behavior*, *67*, 207–220. https://doi.org/10.1016/j.chb.2016.10.026

Goodhue, D. L., & Thompson, R. L. (1995). Task-Technology Fit and Individual Performance. *MIS Quarterly*, *19*(2), 213–236. https://doi.org/10.2307/249689

Goodyear, P. (2015). Teaching as design. *HERDSA Review of Higher Education*, *2*, 27–50.

Google Scholar. (2016). Educational Technology - Google Scholar Metrics. Retrieved January 15, 2017, from https://scholar.google.ca/citations?view_op=top_venues&hl=en&vq=soc_educationaltechnology

Granovetter, M. (1982). The strength of weak ties: A network theory revisited. In P. V. Marsden & N. Lin (Eds.), *Social structure and network analysis* (pp. 105–130). Beverly Hill, CA: Sage.

Greller, W., & Drachsler, H. (2012). Translating learning into numbers: A generic framework for learning analytics. *Journal of Educational Technology & Society*, *15*(3), 42–57.

Hattie, J. (2012). *Visible Learning for Teachers: Maximizing Impact on Learning*. Routledge.

Jayaprakash, S. M., Moody, E. W., Lauría, E. J. M., 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.

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. In *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge* (pp. 314–323). New York, NY, USA: ACM. https://doi.org/10.1145/2883851.2883928

Käser, T., Klingler, S., & Gross, M. H. (2016). When to stop?: towards universal instructional policies. In *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge, LAK 2016, Edinburgh, United Kingdom, April 25-29, 2016* (pp. 289–298). https://doi.org/10.1145/2883851.2883961

Kovanović, V., Gašević, D., Joksimović, S., Hatala, M., & Adesope, O. (2015). Analytics of communities of inquiry: Effects of learning technology use on cognitive presence in asynchronous online discussions. *The Internet and Higher Education*, *27*, 74–89. https://doi.org/10.1016/j.iheduc.2015.06.002

Kovanović, V., Joksimović, S., Waters, Z., Gašević, D., Kitto, K., Hatala, M., & Siemens, G. (2016). Towards Automated Content Analysis of Discussion Transcripts: A Cognitive Presence Case. In *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge* (pp. 15–24). New York, NY, USA: ACM. https://doi.org/10.1145/2883851.2883950

Krumm, A. E., Waddington, R. J., Teasley, S. D., & Lonn, S. (2014). A Learning Management System-Based Early Warning System for Academic Advising in Undergraduate Engineering. In J. A. Larusson & B. White (Eds.), *Learning Analytics* (pp. 103–119). Springer New York. Retrieved from http://link.springer.com/chapter/10.1007/978-1-4614-3305-7_6

Liu, Z., Nersessian, N. J., & Stasko, J. T. (2008). Distributed cognition as a theoretical framework for information visualization. *IEEE Transactions on Visualization and Computer Graphics*, *14*(6), 1173–1180.

Lockyer, L., Heathcote, E., & Dawson, S. (2013). Informing Pedagogical Action Aligning Learning Analytics With Learning Design. *American Behavioral Scientist*, *57*(10), 1439–1459. https://doi.org/10.1177/0002764213479367

Long, P. D., Siemens, G., Conole, G., & Gašević, D. (Eds.). (2011). *Proceedings of the 1st International Conference on Learning Analytics and Knowledge (LAK'11)*. New York, NY, USA: ACM.

Lusher, D., Koskinen, J., & Robins, G. (2012). *Exponential Random Graph Models for Social Networks: Theory, Methods, and Applications*. Cambridge University Press.

Lust, G., Elen, J., & Clarebout, G. (2013). Students' tool-use within a web enhanced course: Explanatory mechanisms of students' tool-use pattern. *Computers in Human Behavior*, *29*(5), 2013–2021. https://doi.org/10.1016/j.chb.2013.03.014

Macfadyen, L. P., & Dawson, S. (2012). Numbers Are Not Enough. Why e-Learning Analytics Failed to Inform an Institutional Strategic Plan. *Educational Technology & Society*, *15*(3).

Marbouti, F., & Wise, A. F. (2016). Starburst: a new graphical interface to support purposeful attention to others' posts in online discussions. *Educational Technology Research and Development*, *64*(1), 87–113. https://doi.org/10.1007/s11423-015-9400-y

Molenaar, I. (2014). Advances in temporal analysis in learning and instruction. *Frontline Learning Research*, *2*(4), 15–24. https://doi.org/10.14786/flr.v2i4.118

Pechenizkiy, M., & Gašević, D. (2014). Introduction into Sparks of the Learning Analytics Future. *Journal of Learning Analytics*, *1*(3), 145–149. https://doi.org/10.18608/jla.2014.13.8

Pijeira-Díaz, H. J., Drachsler, H., Järvelä, S., & Kirschner, P. A. (2016). Investigating Collaborative Learning Success with Physiological Coupling Indices Based on Electrodermal Activity. In *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge* (pp. 64–73). New York, NY, USA: ACM. https://doi.org/10.1145/2883851.2883897

Reimann, P. (2016). Connecting learning analytics with learning research: the role of design-based research. *Learning: Research and Practice*, *2*(2), 130–142. https://doi.org/10.1080/23735082.2016.1210198

Rogers, T., Gašević, D., & Dawson, S. (2016). Learning analytics and the imperative for theory driven research. In C. Haythornthwaite, R. Andrews, J. Fransma, & E. Meyers (Eds.), *The SAGE Handbook of E-Learning Research, 2nd edition* (pp. 232–250). London, UK: SAGE Publications Ltd.

Roll, I., & Winne, P. H. (2015). Understanding, evaluating, and supporting self-regulated learning using learning analytics. *Journal of Learning Analytics*, *2*(1), 7–12.

Romero, C., & Ventura, S. (2010). Educational Data Mining: A Review of the State of the Art. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, *40*(6), 601–618. https://doi.org/10.1109/TSMCC.2010.2053532

Scheffel, M., Drachsler, H., Stoyanov, S., & Specht, M. (2014). Quality Indicators for Learning Analytics. *Educational Technology & Society*, *17*(4), 117–132.

Siadaty, M., Gašević, D., & Hatala, M. (2016). Trace-based Micro-analytic Measurement of Self-Regulated Learning Processes. *Journal of Learning Analytics*, *3*(1), 183–214. https://doi.org/10.18608/jla.2016.31.11

Siemens, G., & Baker, R. S. J. d. (2012). Learning Analytics and Educational Data Mining: Towards Communication and Collaboration. In *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge* (pp. 252–254). New York, NY, USA: ACM. https://doi.org/10.1145/2330601.2330661

Siemens, G., & Gasevic, D. (2012). Guest Editorial-Learning and Knowledge Analytics. *Educational Technology & Society*, *15*(3), 1–2.

Slater, S., Joksimović, S., Kovanovic, V., Baker, R. S., & Gasevic, D. (in press). Tools for Educational Data Mining: A Review. *Journal of Educational and Behavioral Statistics*. https://doi.org/10.3102/1076998616666808

Steiner, C., Kickmeier-Rust, M., & Türker, M. A. (2014). *Review Article about LA and EDM Approaches* (No. Deliverable D3.1). European Commission LEA's Box Project. Retrieved from http://css-kmi.tugraz.at/mkrwww/leas-box/downloads/D3.1.pdf

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction*, *4*(4), 295–312. https://doi.org/10.1016/0959-4752(94)90003-5

Tanes, Z., Arnold, K. E., King, A. S., & Remnet, M. A. (2011). Using Signals for appropriate feedback: Perceptions and practices. *Computers & Education*, *57*(4), 2414–2422. https://doi.org/10.1016/j.compedu.2011.05.016

Tsai, Y.-S., & Gašević, D. (in press). Learning analytics in higher education – challenges and policies. In *Proceedings of the 7th International Conference on Learning Analytics and Knowledge (LAK 2017)*. ACM Press.

Verbert, K., Duval, E., Klerkx, J., Govaerts, S., & Santos, J. L. (2013). Learning Analytics Dashboard Applications. *American Behavioral Scientist*, *57*(10), 1500–1509. https://doi.org/10.1177/0002764213479363

Vessey, I. (1991). Cognitive Fit: A Theory-Based Analysis of the Graphs Versus Tables Literature*. *Decision Sciences*, *22*(2), 219–240. https://doi.org/10.1111/j.1540-5915.1991.tb00344.x

Wasserman, S. (1994). *Social network analysis: Methods and applications* (Vol. 8). Cambridge university press.

Wasson, B., & Hansen, C. (2016). Data literacy and use for teaching. In P. Reimann, S. Bull, M. Kickmeier-Rust, R. Vatrapu, & B. Wasson (Eds.), *Measuring and Visualizing Learning in the Information-Rich Classroom* (pp. 56–73). New York: Routledge.

Winne, P. H. (1997). Experimenting to bootstrap self-regulated learning. *Journal of Educational Psychology*, *89*(3), 397–410. https://doi.org/10.1037/0022-0663.89.3.397

Winne, P. H. (2006). How Software Technologies Can Improve Research on Learning and Bolster School Reform. *Educational Psychologist*, *41*(1), 5–17. https://doi.org/10.1207/s15326985ep4101_3

Winne, P. H., & Hadwin, A. F. (1998). Studying as self-regulated learning. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 277–304). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

Wise, A. (2014). Designing Pedagogical Interventions to Support Student Use of Learning Analytics. In *Proceedings of the Fourth International Conference on Learning Analytics And Knowledge* (pp. 203–211). New York, NY, USA: ACM. https://doi.org/10.1145/2567574.2567588

Wise, A., & Shaffer, D. W. (2015). Why Theory Matters More than Ever in the Age of Big Data. *Journal of Learning Analytics*, *2*(2), 5–13. https://doi.org/10.18608/jla.2015.22.2

Wise, A., Zhao, Y., & Hausknecht, S. (2014). Learning Analytics for Online Discussions: Embedded and Extracted Approaches. *Journal of Learning Analytics*, *1*(2), 48–71.

Zhou, M., & Winne, P. H. (2012). Modeling academic achievement by self-reported versus traced goal orientation. *Learning and Instruction*, *22*(6), 413–419. https://doi.org/10.1016/j.learninstruc.2012.03.004