Educational Data Mining and Learning Analytics for Improving Online Learning Environments

Yancy Vance Paredes, Robert F. Siegle, I-Han Hsiao, and Scotty D. Craig
Arizona State University
Mesa, AZ

The proliferation of educational technology systems has led to the advent of a large number of datasets related to learner interaction. New fields have emerged which aim to use this data to identify interventions that could help the learners become efficient and effective in their learning. However, these systems have to follow user-centered design principles to ensure that the system is usable and the data is of high quality. Human factors literature is limited on the topics regarding Educational Data Mining (EDM) and Learning Analytics (LA). To develop improved educational systems, it is important for human factors engineers to be exposed to these data-oriented fields. This paper aims to provide a brief introduction to the fields of EDM and LA, discuss data visualization and dashboards that are used to convey results to learners, and finally to identify where human factors can aid other fields.

INTRODUCTION

The increasing accessibility of computers has spurred enormous growth in the development of educational technology systems. These systems are capable of capturing a wide range of data about the user including behavior, interaction, and performance. The ability for these systems to capture fine-grained data about the users at greater speeds than ever before has brought about the concept of big data.

Big data is often characterized by volume, velocity, and variety which are known as the three V's (Laney, 2001). Volume refers to the huge size of the data, velocity as the data are being recorded at high speeds, and variety as various types are being captured. Traditional methods that require manual processing of data are being overshadowed by newer approaches that can process the vast amounts of data in semi or fully automated ways (Carnahan et al., 2013). This emerging concept of big data gave birth to new fields that are attempting to make sense of or find patterns in enormous datasets that a simple analysis would not be able to uncover.

Two emerging fields that focus on big data are Educational Data Mining (EDM) and Learning Analytics (LA). As these fields mature, issues are being raised regarding the design of the educational systems. Although these fields emphasize on the analysis of the data, the design of educational systems could affect the quality of the data and must be taken into account.

As human factors engineers, it is important to apply user-centered design principles when designing these educational systems. However, a review of the literature yields a limited amount of discussions concerning design principles in the context of educational systems. There have been ongoing efforts and conversations to introduce related fields such as data science and machine learning (Ma & Drury, 2003; Lau et al., 2018; Hannon et al., 2019). Therefore, this paper aims to continue the conversation and provide an introduction to EDM and LA.

EDUCATIONAL DATA MINING

EDM is primarily concerned with the development, research, and application of computerized techniques used to analyze large amounts of educational data (i.e., captured from educational settings) with the hopes of detecting meaningful patterns (Romero & Ventura, 2013). This interdisciplinary field sits at the intersection of Computer Science, Education, and Statistics. EDM research aims to create a better understanding of how students learn and to identify the best settings for learning (Romero & Ventura, 2013), and to turn students into more effective learners (Baker, 2013). In addition, improving learning has been one of the core objectives of the field. However, such measures are not easily obtainable which is why improvements in performance have been used to estimate this instead.

Romero and Ventura (2013) have shown a wide variety of data types are being analyzed using EDM techniques. There have been studies that analyze interaction data from an individual learner as the user interacts with the system, or from collaborating learners as they interact with one another. The data can also be administrative in nature coming from the school or the teacher. Oftentimes, demographics data such as age and sex are utilized. Other research has used emotional states or student affectivity data. All these different types of data are collected either from the traditional or the computer-based education (e.g., learning management systems, intelligent tutoring systems) environments and have typical characteristics such as multiple levels of hierarchy, context, fine-grained, and longitudinal.

Topics of interest among the EDM research community center around developing generic frameworks and methods. There is also interest in building systems that use data for adaptation and personalization of education. Identifying best practices and improving how to better support teachers are key areas being examined by the EDM community.

Common Methods in EDM

Several EDM techniques have origins that can be traced to data mining. Romero and Ventura (2013) have outlined several common methods used in EDM. Table 1 below provides a brief summary of these methods.

Table 1
Common EDM Methods

Method	Description
Prediction	Infer a target attribute
Clustering	Identifying data points that are similar
Outlier Detection	Identifying data points that are significantly different
Relationship Mining	Determining the relationship between variables
Social Network Analysis	Measure relationships among entities in a networked context
Process Mining	Using event logs to come up with a visual representation of the whole process
Text Mining	Extracting useful information from text data
Distillation of Data for Human Judgement	Representing data in a more comprehensible way
Discovery with Models	Use of previously validated models in another analysis
Knowledge Tracing	Estimation of a student's master on a particular skill
Nonnegative Matrix Factorization	Technique that enables interpretation in terms of Q-Matrix in a straightforward manner

Applications of EDM

Since its inception, predicting students' performance has been the most popular application of EDM techniques (Romero & Ventura, 2013). There has also been interest in modeling the user by developing and tuning cognitive models to represent their skills and declarative knowledge (Frias-Martinez et al., 2006). Additionally, instructors and developers of learning content have seen the benefit in the utilization of EDM for constructing coursewares (Garcia et al., 2009). Lastly, using the vast data available, parameters to probabilistic models can be inferred (also known as parameter estimation) to determine the probability of an event of interest to occur (Wauters et al., 2011).

LEARNING ANALYTICS

LA encompasses the measurement, collection, analysis, and reporting of data about learners. LA's aim is to understand and optimize both the learning process and the environment in which it occurs (1st International Conference on Learning Analytics and Knowledge, 2010). In fact, LA has been considered a fast-growing area of technology learning research, pioneered by those that envisioned it as an approach to education that is guided by pedagogy (Ferguson, 2012; Greller & Drachsler, 2012). Many factors drive the research on learning analytics (Buckingham-Shum, Gasevic, & Ferguson, 2012). Among these factors are the increased motivation, autonomy, effectiveness, and efficiency of both the learners and the educators.

Papamitsiou and Economides (2014) were able to identify common learning settings in the literature of LA research. This includes virtual learning environments, learning management systems, cognitive tutors, class-based and web-based environments, and mobile settings. They further highlight how recent studies have started exploring massive open online courses (MOOCs) and social learning platforms. This widespread use of online learning environments led to the explosion of big data (e.g., interaction data, personal and academic information), which learning analytics aims to exploit the potential of (Ferguson, 2012). Classification, clustering, and regression (both logistic and multiple) have been identified as popular techniques in LA (Papamitsiou & Economides, 2014). There has been growing interest in the use of discovery with models approach recently.

One growing area in the field of LA is multimodal learning analytics (MMLA). With the increasing number of technologies that are capable of collecting learner artifacts, new insights into the learner's learning trajectories can be explored (Blikstein & Worsley, 2016). Experts are looking to expand the LA techniques beyond the traditional log-files (Ochoa, 2017). Instruments such as wearable cameras, biosensors, and eye trackers have been widely used to track multiple human activities. These multiple sources of user information could be integrated with current data collection methods to evaluate the complex cognitive abilities of the learners. Although promising, one issue on MMLA is its impact on learning. The positive impact on learning should be large enough to compensate for the high complexity involved in the acquisition and analysis of the data (Ochoa, 2017).

Objectives

The literature on LA has a wide span of research goals and objectives (Papamitsiou & Economides, 2014), the most common being student or student behavior modeling. These studies examine how to detect, identify, and model the learning behavior of students in different learning environments. Another objective is to predict student performance. This involves exploring, identifying, and evaluating factors that affect performance. LA has also been used to increase the self-reflection and self-awareness of students through the use of visualizations that inform them of their progress and performance. Due to the vast amount of data captured by systems, especially in large systems such as MOOCs, predicting the dropout and retention of learners has become popular with the intention to provide early interventions could be provided to students. Finally, other research has aimed to use data known to the system for improving feedback and assessment services provided to students through meaningful feedback and the recommendation of resources.

Case Studies

Course Signals is a prominent example of a system that utilized the power of learning analytics. It is a system developed at Purdue University which allows instructors to give real-time feedback to students through the use of faculty dashboards (Arnold & Pistilli, 2012). Predictive models, which run on-demand, use the vast data captured by multiple systems in the university. The university saw significantly higher retention rates on students who have used the system at least once over those who have not used it at all.

Weaknesses

Ferguson (2012) identified some of the future challenges in learning analytics. Currently, research that focuses on cognition, metacognition, and pedagogy are underrepresented. Complex datasets outside the formal learning environment need to be explored if learning environments are to be fully optimized. Most importantly, a clear set of ethical guidelines has to be developed and applied on data privacy.

Learning Analytics and Educational Data Mining

The research areas of LA and EDM have some similarities. However, they are two separate fields. As these two fields start to mature, their central themes become clearer (Ferguson, 2012). LA focuses on the educational challenge: *How can we optimize opportunities for online learning?* EDM focuses on the technical challenge: *How can we extract value from these big sets of learning-related data?* Siemens and Baker (2012) identified five different aspects they used to distinguish one from the other: (1) the type of discovery being prioritized, (2) the reductionist and holistic frameworks, (3) origins, (4) adaptation and personalization, and (5) popular techniques and methods used.

Although both fields use visualizations and some other methods to automate the discovery process, LA puts a greater focus on leveraging human judgment and uses automated discovery as a way to inform humans in decision-making. EDM places the focus on automated discovery and seeks human judgment (e.g., experts) in the form of providing labels for classification.

In the field of LA, systems are viewed holistically to understand them as a whole in their full complexity. In EDM, systems are reduced into their multiple components with each component being analyzed to understand the relationship and interaction among them.

The origins of the LA can be traced from the fields of semantic web, intelligent curriculum, outcome prediction, and system interventions. EDM's origins are from the fields of educational software and student modeling, particularly those in predicting course outcomes.

Models produced in LA are mostly designed to empower various stakeholders (e.g., instructors and students) for them to make informed decisions. The models produced in EDM are mainly designed with adaptation in mind. These models are then used to automate systems that do not have humans in the loop to adapt (e.g., intelligent tutoring systems).

Finally, techniques that are popular in LA include social network analysis, sentiment analysis, influence analytics, discourse analysis, learner success prediction, concept analysis, and sensemaking models. EDM typically uses classification, clustering, Bayesian modeling, relationship mining, discovery with models, and visualization.

DATA VISUALIZATION

To be able to make sense of the vast amount of learning data available, two main approaches to LA can be employed (Ruiperez-Valiente et al., 2014). Systems such as intelligent tutoring systems or recommender systems can be developed to automatically process the data. Another approach is through direct visualization to the stakeholders. This presupposes the ability of humans to recognize or discover patterns from visualization (e.g., trends, outliers, clusters, gap). Cognition is further amplified when interactive elements are utilized in information visualizations since this facilitates exploratory data analysis (Card et al., 1999).

Exploratory Data Analysis

The process of visual data exploration, also known as hypothesis generation, comes into play when the humans are taken into account in the data mining process. Using the interactive elements in the visualization new insights can be formed and conclusions can be drawn. Shneiderman's (1996) Information Seeking Mantra or the "overview first, zoom and filter, then details-on-demand" (p. 337) is one popular visual exploration paradigm that serves as a guideline on how to design effective visualizations. Users must be able to see an overview of the visualization to look for interesting patterns. Afterward, they should be able to drill down and access the details of the data for them to analyze the patterns. Recent studies suggest keeping the overview within view while the subset is focused on using a different visualization technique (e.g., distortion) (Keim, 2002). Several data types can be visualized which include 1-dimensional, 2-dimensional, 3-dimensional, temporal, multi-dimensional, tree, and network (Shneiderman, 1966). Some of the techniques that can be employed to visualize data include standard 2D, 3D, geometrically transformed displays, iconic displays, dense pixel displays, and stack displays. The following techniques used for interaction and distortion are dynamic projects, interactive filtering, interactive zooming, interactive distortion, and interactive linking and brushing.

Designing Information Visualization Systems

Designing and developing visualization systems involve a series of steps. Klerkx and colleagues (2017) outlined a guideline which has six steps. The first step is to understand your goal and determine why the visualization is needed and to whom it is intended. Identifying how your goals can be achieved is crucial in this step. The next step is to acquire and pre-process your data. This includes cleaning the data in which experts in the field note take about 80% of the time and effort of the entire process. Data that are irrelevant or those that do not help answer the main question are filtered out. Once the data is ready, the next step is to map the data to an appropriate design that would best represent the data and befitting for the target audience. The next steps involve documenting the process by providing a writeup of the rationale for the decisions made that led to the final design and noting any alternatives that could have possibly been chosen.

How the visualization evolved from the initial state to the current step should be discussed. Since the process of visual analysis involves an iteration of view creation, exploration, and refinement, the next step is to add interaction techniques. This could be in the form of brushing and linking, histogram sliders, zoomable maps, and dynamic query filter widgets. Finally, the last step is to evaluate continuously by taking in possible feedback from the users to further improve the system.

Dashboards

One prominent technique for information visualization is through the use of dashboards. Dashboards are commonly used to display the most important information to users by consolidating information on a single computer screen that can be seen at a glance to achieve one or more objectives (Few, 2013). The data used for visualization are lifted from patterns that emerged from the big data. In the context of education, such dashboards are commonly referred to as educational dashboards, learning dashboards, or learning analytics dashboards. These dashboards are used to intuitively display results of EDM with the aim of supporting the learning of students and the improvement of their performance (Yoo et al., 2015). This supports learning and teaching by visualizing learning traces for learners and teachers (Verbert et al., 2013), and by providing a current and historical state of the learners (Few, 2006). The area of educational dashboards is still new and lacks a set of principles of the field (Yoo et al., 2015).

Many types of information could be incorporated into learning dashboards (Verbert et al., 2014). These types include artifacts produced by learners such as blog posts or items that end up in students' project portfolio. Others are social interactions that include face-to-face, group or blog comments. How students use resources (e.g., watching a video) is also information that can be used. Another widely used information type is the amount of time spent on a task which can be used by teachers to identify if a student is at-risk and by students to compare their efforts among their peers. Typical information about test and self-assessment results are also used to indicate the learning progress of the students.

The first step in evaluating a dashboard is identifying its intended goals. The next is to identify its impact on learners' affect and motivation. Afterward, identify the system's usability which is not limited to determining whether it is useful or not. Instead, the ability of the users to trust the dashboard (i.e., whether the users agree with what the dashboard is presenting to them) should be examined and how the users interpreted feedback should be assessed as well. In terms of evaluating the effectiveness of dashboards, Jivet and colleagues (2018) suggest that data triangulation of self-reported data, tracked data, and assessment data must be performed to validate its effects. Only validated instruments should be used to assess the impacts of the dashboard on learners.

DISCUSSION

Ferguson (2012) notes the need to work with complex data that are beyond the traditional learning environment (e.g., biometric data, mobile data, mood data). However, complex data is challenging to capture as the use of an external sensor would require the system to have a different format. This belabors the data collection and cleaning steps. The data being collected revolves around the experiences of the learner. The Advanced Distributed Learning Initiative developed a standardized way to capture these learning experiences in the form of Experience API (xAPI) also known as Tin Can API. xAPI are statements capable of describing learning activities that can be shared across different systems (Kevan & Sorensen Irvine, 2017). Each statement consists of three parts: an actor, a verb, and an object. Contextual information can be added to provide more details on the learning activity. Murphy and colleagues (2016) found that in the context of training technology, xAPI is able to capture and share human performance data. Using a standardized approach in data collection would allow standardized analysis toolkits to be developed, which would significantly reduce the effort in data cleaning. There are ongoing efforts in developing toolkits that process xAPI statements and apply LA algorithms (Yet Analytics, 2019). However, despite the potential offered by xAPI research on the use of xAPI is still limited due to the slow adoption of providers.

One concern that is often raised in the fields of LA and EDM is the issue of ethics and privacy. Since the data being collected contains student information, researchers are cautious in how they handle this data. To address this, several frameworks have been proposed to outline how to ethically perform EDM or LA on educational data (Slade & Prinsloo, 2013; Pardo & Siemens, 2014; Drachsler & Greller, 2016).

Finally, as human factors engineers are exposed to the data-oriented fields of LA and EDM, they become equipped with knowledge of designing and improving educational systems. Techniques from LA and EDM could be utilized to supplement the traditional data collection practices of the human factors field as the data becomes more complex. For example, when performing usability testing, additional information from the system logs (e.g., time on task, errors, or process flows) allows the designers to become aware of where the users spend most of their time in the system. This aids designers in identifying which parts of the system have potential problems. It also provides an automated way of capturing how users use the system. These supplementary data can help developers determine whether difficulties experienced by the users are caused by the user's lack of understanding or by the system itself. This additional dimension in investigating the system could lead to the creation of better educational systems and improvements that would lead students to success.

ACKNOWLEDGMENTS

The current paper was partially funded by HQ0034-19-C-0015, a contract from the United States Department of Defense Advanced Distributed Learning Initiative. The statements within this work do not reflect the opinion, view or policy of the funding agency.

REFERENCES

1st International Conference on Learning Analytics and Knowledge. (2010). Retrieved October 3, 2019, from https://tekri.athabascau.ca/analytics/.

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). ACM.

Baker, R. J. d. (2013). Learning, schooling, and data analytics. In M. Murphy, S. Redding, & J. Twyman (Eds.), *Handbook on innovations in learning* (pp. 179-190). Philadelphia, PA: Center on Innovations in Learning, Temple University; Charlotte, NC: Information Age Publishing.

Blikstein, P., & Worsley, M. (2016). Multimodal Learning Analytics and Education Data Mining: Using computational technologies to measure complex learning tasks. *Journal of Learning Analytics, 3*(2), 220-238.

Buckingham-Shum, S., Gasevic, D., & Ferguson, R. (Eds.). (2012). *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge.* ACM.

Card, S., Mackinlay, J., & Shneiderman, B. (1999). *Readings in information visualization: Using vision to think*. San Francisco, CA: Morgan Kaufmann.

Carnahan, B., Meyer, G., & Kuntz, L.-A. (2003). Comparing statistical and machine learning classifiers: Alternatives for predictive modeling in human factors research. *Human Factors*, *45*(3), 408–423.

Drachsler, H., & Greller, W. (2016). Privacy and analytics: It's a DELICATE issue a checklist for trusted learning analytics. In *Proceedings of the 6th International Conference on Learning Analytics and Knowledge* (pp. 89-98). ACM.

Ferguson, R. (2012). Learning analytics: Drivers, developments and challenges. *International Journal of Technology Enhanced Learning*, *4*(5/6), 304-317.

Few, S. (2006). *Information dashboard design: The effective visual communication of data* (1st ed.). Cambridge, MA: O'Reilly.

Few, S. (2013). *Information dashboard design: Displaying data for at-a-glance monitoring* (Vol. 5). Burlingame, CA: Analytics Press.

Frias-Martinez, E., Chen, S. Y., & Liu, X. (2006). Survey of data mining approaches to user modeling for adaptive hypermedia. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, *36*(6), 734-749.

Garcia, E., Romero, C., Ventura, S., Gea, M., & De Castro, C. (2009). Collaborative Data Mining Tool for Education. *International Working Group on Educational Data Mining*.

Greller, W., & Drachsler, H. (2012). Translating Learning into Numbers: A Generic Framework for Learning Analytics. *Journal of Educational Technology & Society*, *15*(3), 42–57.

Hannon, D., Rantanen, E., Sawyer, B., Ptucha, R., Hughes, A., Darveau, K., & Lee, J. D. (2019). A Human Factors Engineering Education Perspective on Data Science, Machine Learning and Automation. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *63*(1), 488–492.

Jivet, I., Scheffel, M., Specht, M., & Drachsler, H. (2018). License to evaluate: Preparing learning analytics dashboards for educational practice. In *Proceedings of the 8th International Conference on Learning Analytics and Knowledge* (pp. 31-40). ACM.

Keim, D. A. (2002). Information visualization and visual data mining. *IEEE Transactions on Visualization and Computer Graphics*, *8*(1), 1-8.

Kevan, J. M., & Sorensen Irvine, C. K. (2017). Academic technology for competency-based education in higher education. In K. Rasmussen, P. Northrup, & R. Colson (Eds.), *Handbook of research on competency-based education in university settings* (pp. 116-134). Hershey, PA: IGI Global.

Klerkx, J., Verbert, K., and Duval, E. (2017). Learning Analytics Dashboards. In C. Lang, G. Siemens, A. F. Wise & D. Gasevic (Eds.), *The Handbook of Learning Analytics*. (1st ed., pp. 143-150).

Laney, D. (2001). 3D data management: Controlling data volume, velocity and variety. *META group research note*, *6*(70), 1.

Lau, N., Fridman, L., Borghetti, B. J., & Lee, J. D. (2018). Machine Learning and Human Factors: Status, Applications, and Future Directions. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *62*(1), 135–138.

Ma, J., & Drury, C. G. (2003). The Human Factors Issues in Data Mining. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *47*(4), 716–720.

Murphy, J., Hannigan, F., Hruska, M., Medford, A., & Diaz, G. (2016). Leveraging interoperable data to improve training effectiveness using the Experience API (xAPI). In *International Conference on Augmented Cognition* (pp. 46-54). Springer, Cham.

Ochoa, X. (2017). Multimodal Learning Analytics. In C. Lang, G. Siemens, A. F. Wise & D. Gasevic (Eds.), *The Handbook of Learning Analytics*. (1st ed., pp. 129-141).

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

Pardo, A., & Siemens, G. (2014). Ethical and privacy principles for learning analytics. *British Journal of Educational Technology*, *45*(3), 438-450.

Romero, C., & Ventura, S. (2013). Data mining in education. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *3*(1), 12-27.

Ruipérez-Valiente, J. A., Muñoz-Merino, P. J., Leony, D., & Kloos, C. D. (2015). ALAS-KA: A learning analytics extension for better understanding the learning process in the Khan Academy platform. *Computers in Human Behavior*, *47*, 139-148.

Shneiderman, B. (1996). The eyes have it: A task by data type taxonomy for information visualizations. In *Proceedings 1996 IEEE Symposium on Visual Languages* (pp. 336-343). IEEE.

Siemens, G., & d Baker, R. S. (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). ACM.

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

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

Verbert, K., Govaerts, S., Duval, E., Santos, J. L., Assche, F., Parra, G., & Klerkx, J. (2014). Learning dashboards: An overview and future research opportunities. *Personal and Ubiquitous Computing*, *18*(6), 1499-1514.

Wauters, K., Desmet, P., & Van Den Noortgate, W. (2011). Acquiring item difficulty estimates: A collaborative effort of data and judgment. In *Proceedings of the 4th International Conference on Educational Data Mining* (pp. 121-127). Eindhoven University of Technology.

Yet Analytics. (2019). *Data analytics and visualization environment for xAPI and the total learning architecture: DAVE learning analytics algorithms*. ADL.

Yoo, Y., Lee, H., Jo, I. H., & Park, Y. (2015). Educational dashboards for smart learning: Review of case studies. In *Emerging Issues in Smart Learning* (pp. 145-155). Springer, Berlin, Heidelberg.

REPORT DOCUMENTATION PAGE			*Form Approved OMB No. 0704-0188*

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From - To)*
10/5/2020	Proceedings Paper	2020-2021

4. TITLE AND SUBTITLE
Educational Data Mining and Learning Analytics for Improving Online Learning Environments

5a. CONTRACT NUMBER
HQ0034-19-C-0015

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
0603769D8Z

6. AUTHOR(S)
Yancy Vance Paredes, Robert F. Siegle, IHan Hsiao, and Scotty D. Craig

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
ARIZONA STATE UNIVERSITY ORSPA KELBRINA DAVIS 660 S MILL AVE STE 312 TEMPE AZ 85281-3670

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
OUSD Personnel & Readiness
Advanced Distributed Learning Initiative
13501 Ingenuity Drive, Suite 248
Orlando, Florida 32826

10. SPONSOR/MONITOR'S ACRONYM(S)
OUSD/P&R/FE&T/ADLI

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Distribution A

13. SUPPLEMENTARY NOTES

14. ABSTRACT
The proliferation of educational technology systems has led to the advent of a large number of datasets related to learner interaction. New fields have emerged which aim to use this data to identify interventions that could help the learners become efficient and effective in their learning. However, these systems have to follow user-centered design principles to ensure that the system is usable and the data is of high quality. Human factors literature is limited on the topics regarding Educational Data Mining (EDM) and Learning Analytics (LA). To develop improved educational systems, it is important for human factors engineers to be exposed to these data-oriented fields. This paper aims to provide a brief introduction to the fields of EDM and LA, discuss data visualization and dashboards that are used to convey results to learners, and finally to identify where human factors can aid other fields.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	c. THIS PAGE	**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
U	U	U	UU	5	Preston Bush
					19b. TELEPHONE NUMBER *(Include area code)* 571-480-4640 Ext. 4659