Learning Analytics Education: A case study, review of current programs, and recommendations for instructors

René F. Kizilcec

Bowers College of Computing and Information Science, Cornell University
208 Gates Hall, 107 Hoy Road, Ithaca, NY 14853, USA
kizilcec@cornell.edu

Dan Davis

Georgetown University
3520 Prospect ST. NW, Washington, DC 20057
dan7davis@gmail.com

Introduction

The interdisciplinary field of learning analytics emerged in 2008 and quickly grew into a global community of researchers, practitioners, and educators who have made important scientific and applied contributions (Clow, 2013; Siemens, 2013). Journals, conferences, workshops, and informal online outlets such as blogs have served as venues for knowledge exchange, co-creation, and inspiration. As the field matures, institutions of higher education increasingly offer courses, certificates, and degree programs in learning analytics to disseminate the theories, methods, applications, and values of this field. These educational programs help train the next generations of leaders in learning analytics research, practice, and policy. They also encourage more people to work in areas related to learning analytics, especially those looking to combine an interest in data science and technology with a desire to effect positive change in society. These efforts to teach and learn learning analytics in formal and informal educational environments are the focus of this chapter. We begin with a survey of the landscape of current learning analytics programs and examine what topics and pedagogies are represented. This is followed by an in-depth case study of a learning analytics course offered to undergraduate and graduate students at Cornell University. The case study demonstrates a pedagogical approach to learning analytics education for students with a more technical emphasis. Finally, we discuss the current state of learning analytics education and identify challenges and opportunities for learning analytics education going forward. This chapter contributes to practicable learning analytics by providing evidence on the status quo of teaching and learning learning analytics with a comprehensive review of current learning analytics programs and a case study of a university course, and by offering a set of actionable guidelines for the community to consider when designing learning analytics courses.

Learning analytics education has a wide range of audiences and objectives. Students, teachers, instructional designers, parents, professional student advisers, and school leaders are increasingly likely to interact with or be affected by learning analytics models and applications. They can benefit from understanding the assumptions, data inputs, engineering and design choices underlying these models and applications. It helps them make informed judgments about the relevance and appropriateness of different learning analytics for their use case and the kinds of inferences they can draw from the information to inform their actions and policy decisions. There are also important

audiences outside of traditional formal education environments with a stake in learning analytics education. The growth of interest in lifelong learning and demand for continuous skill development in the labor market has elevated the role of professional development. Working learners need to make decisions year after year about which formal or informal educational opportunities to pursue and whether they are effectively learning the knowledge and skills they need. Human resource departments, which tend to oversee professional development programs and policies, need to make informed decisions about which learning opportunities to offer or incentivize, and how to evaluate employees' learning outcomes and their downstream effects on performance at the intersection between learning analytics and people analytics (Tursunbayeva, Di Lauro and Pagliari, 2018). In some high-stakes work environments, such as aviation, medical, and military contexts, precise training and assessment analytics have already been in use and other work environments are eager to adopt a targeted approach to professional development with learning analytics. Given this wide range of audiences with varying objectives for learning about learning analytics, there is not just one right learning analytics curriculum for everyone as illustrated in our survey of programs and the case study.

The field of learning analytics keeps evolving, building on expertise from various scientific disciplines, and its applications are integrated into more and more real-world contexts with different domain-specific knowledge and skills. Learning analytics is grounded in the learning sciences, including cognitive science, social and educational psychology (Sawyer, 2005), and in the computational social sciences, including computer and data science, network analysis, data visualization, and statistics (Lazer *et al.*, 2009). Learning analytics research and practice relies on combinations of theory and methodology from these two clusters of disciplines. Early adoptions of learning analytics applications prompted questions about ethics and privacy, which has started to bring in disciplinary expertise from law, sociology, public policy, and critical studies. Moreover, domain experts are frequently involved in domain-specific learning analytics to provide context and address particular issues in that domain. Altogether there is a diversity of disciplinary backgrounds represented and engaged in collaborations in learning analytics events and organizations; for example, the Educational Data Mining Society and its International Conference on Educational Data Mining (EDM; started in 2008), the Society of Learning Analytics Research (SoLAR) and its International Conference on Learning Analytics and Knowledge (LAK; started in 2011), and at the ACM Conference on Learning at Scale (L@S; started in 2014). The interdisciplinary nature of learning analytics suggests that a curriculum for learning analytics can be offered by various departments and organizations, not only schools of education. This point is illustrated both in our case study course which is offered by the College of Computing and Information Science and in our survey of the learning analytics education landscape which identifies multiple different departments offering learning analytics programs. The next section provides an overview of educational offerings in the field of learning analytics.

Learning Analytics Education Landscape

We conducted a review to understand the landscape of educational offerings for learning analytics with a focus on the types of programs and institutions offering them. The goal of this survey is to highlight trends in the geography of institutions, disciplinary homes, and types of current learning analytics programs. We used the following methodology to arrive at the list of current programs. Two search strategies were employed to identify relevant programs: (1) Exploratory web searches for "learning analytics curriculum" and "learning analytics [course|workshop|certificate|program]" on Google (English, US) each returned several pages of relevant results. We then screened each result on the first ten pages of search results for relevance and focus on learning analytics, excluding programs that do not focus on learning analytics (e.g., programs about data analytics or about learning science). All relevant programs were added to the list. (2) Targeted web searches for programs at universities that house actively publishing learning analytics researchers, using Google and the university's search page, surfaced additional programs and events, which we screened for relevance and focus to include in the list. Once the list of programs was compiled using these two search methods, we reviewed all available official online materials for each

program (information page, syllabus, timetable, admissions criteria, evaluation criteria, course materials, etc.) to categorize them by program type and record general program information (Table 1). The scope of program characteristics is limited to surface-level information because the amount of openly available program information varies widely across programs. The final list of programs may not be exhaustive or internationally representative due to the nature of Google search in English and socio-cognitive biases of two US-based researchers. Nevertheless, the list provides the first formal overview of the characteristics of currently available—as of February 2022—learning analytics programs that are easily retrievable through English web search.

We observe that there are many different types of learning analytics programs that are offered, including self-paced open educational resource (OER) collections, conference workshops, massive open online courses (MOOCs), university courses, graduate certificates, and even entire master's degree programs. Most of the programs are offered by institutions that are highly ranked globally or nationally, and there is a skew towards programs offered by US universities that charge high tuition fees. However, several programs are international and broadly accessible, such as the OER or MOOC programs, though they do not provide formal university credit for completion. The majority of programs are offered by schools of education or related units, but some programs are offered by schools of information and computer science, or related units. The workload, even for programs of the same type, varies considerably in terms of the number of courses, credit hours, and time allotted for program completion.

This survey of the learning analytics education landscape highlights three major points. First, the field of learning analytics has gained maturity as indicated by high-profile institutions offering dedicated degree programs for learning analytics. More institutions around the world, and especially education schools eager to innovate, may consider this a signal to begin offering learning analytics programs as well. Second, the supply of learning analytics programs is remarkably tailored to diverse learner audiences from college students to graduate students to working professionals, which suggests demand for learning analytics training and credentialing from a broad range of interested parties. And third, the concentration of learning analytics programs in US universities and schools of education may limit global membership and state-of-the-art technology contributions, though there are a number of high-quality OER collections that can facilitate course offerings in more parts of the world and in more disciplinary areas going forward.

In reviewing the available online materials for each program, it quickly became apparent that there is no standard curriculum for learning analytics at this time. While most programs emphasized data literacy and an awareness of common analytic methods and systems as part of their learning goals, there was no common set of topics covered across all programs. Probably the clearest distinction between programs is in terms of how technical their curriculum and assignments are: for example, the seminar course at Georgetown University requires weekly response papers and a research proposal, while the lecture course at Cornell University requires weekly homework projects performing data cleaning and analysis in R. In 2021, SoLAR created an Education Working Group tasked with promoting "the development of high-quality Learning Analytics educational resources" (https://www.solaresearch.org/about/governance/solar-working-groups/). Initiatives from this group have included the development of a public learning analytics dissertation repository and a SoLAR In-Cooperation resource. This initiative invites submissions of any educational project that teaches learning analytics (including, but not limited to, courses, formal or informal programs, and textbooks) to be reviewed by the members of the working group who then provide feedback to ensure quality and consistency of the materials. After addressing the committee's feedback, the project receives an "In-Cooperation with SoLAR" certification, which can be publicly attached to the project to signal its coordination with the learning analytics community. The In-Cooperation project began in 2021 and, at the time of writing, supports the MS in Learning Analytics degree program from the University of Texas Arlington. This type of initiative can also provide guidance to institutions interested in developing new educational programs on learning analytics by recommending a curriculum.

Table 1. Characteristics of current Learning Analytics programs organized by program type.

Type	Level	Institution	Department	Title	Workload	Country
Degree	MA	New York University	School of Culture, Education, and Human Development - Learning Analytics Research Network	Digital Media Design for Learning - Learning Analytics and Educational Data Science (LADS) Specialization	36 credits, 2 years	USA
	MS	Columbia University Teachers College	Graduate School of Education, Health, and Psychology	Learning Analytics	32 credits, 1 year	USA
	MS	University of Texas Arlington	Department of Psychology	Learning Analytics	36 credits, 2.5 years	USA
	MS	University of Wisconsin–Madison	School of Education, Department of Education Psychology	Educational Psychology – Learning Analytics	30 credits, 2 years part-time	USA
	MS	Stanford University	Graduate School of Education	Education Data Science	51 units, 2 years	USA
Certificate	Graduate	Brandeis University	School of Continuing Studies	Graduate Certificate in Learning Analytics	5 courses	USA
	Graduate	Drexel University	School of Education	Learning Analytics Certificate Program	3 courses, 1 year	USA
	Graduate	Monash University	Faculty of Information Technology - Centre for Learning Analytics	Graduate Certificate of Learning Analytics	4 courses, 6 months	Australia
	Graduate	NC State University	Online and Distance Education	Learning Analytics	4 courses, 1-2 years	USA
	Graduate	University of North Dakota	College of Education and Human Development	Learning Analytics Certificate	4 courses, 1 year	USA
	Graduate	University of Florida	College of Education	Learning Analytics	4 courses	USA
	Undergraduate and Graduate	Algonquin College	School of Education	Learning Analytics	3 courses, 1 year	Canada
Course	MOOC	Blackboard	Blackboard Academy	Getting Started with Learning Analytics	4 weeks, 20 hours total	USA
	Graduate	Indiana University	School of Education	Introduction to Educational Data Sciences	3 credits	USA
	Graduate	Norwegian University of Science and Technology	Department of Computer Science	Learning Technology and Analytics	7.5 credits	Norway
	Graduate	KTH Royal Institute of Technology	School of Electrical Engineering and Computer Science	Introduction to Learning Analytics	7.5 credits	Sweden
	Graduate	UMass Amherst	Department of Computer Science	Educational Data Mining and Learner Analytics	3 credits	USA
	Graduate	University of Bergen	Department of Information Science and Media Studies	Introduction to Learning Analytics	7.5 ECTS	Norway

	Graduate	University of South Australia	UniSA Education Futures	Learning Analytics and Digital Learning	4.5 units	Australia
	Graduate	University of Michigan	School of Information	Learning Analytics	3 credits	USA
	Graduate	Georgetown University	Graduate School of Arts and Sciences - Learning Design & Technology Program	Learning Analytics Seminar	3 credits	USA
	MOOC	University of Michigan	via edX	Practical Learning Analytics	4 weeks at 2--8 Hours per week	USA
	MOOC	University of Pennsylvania	via edX	Big Data and Education	8 weeks at 6--12 hours per week	USA
	MOOC	University of Texas Arlington	via edX	Learning Analytics Fundamentals	4 weeks at 5--7 hours per week	USA
	MOOC	—	via Canvas	Learning Analytics and Knowledge 2013	8 weeks, 5-10 hours per week	International
	Undergraduate and Graduate	Cornell University	Department of Information Science	Learning Analytics	3 credits	USA
Workshop	Graduate	Society of Learning Analytics Research	—	Learning Analytics Summer Institute (LASI)	1 week	International
	—	KTH Royal Institute of Technology	—	Nordic Learning Analytics Summer Institute	2 days	Sweden
	—	Society of Learning Analytics Research	International Conference on Learning Analytics & Knowledge	Building the Learning Analytics Curriculum	0.5 days	International
	—	Society of Learning Analytics Research	International Conference on Learning Analytics & Knowledge	Try R: A Gentle intro to R for Learning Analytics	0.5 days	International
	—	University of Eastern Finland	School of Computing	Learning Analytics Summer School	12 days	Finland
	—	Learning Analytics Learning Network	—	Learning Analytics Learning Network (LALN) event hub	—	International
Resource	MOOC	Commonwealth of Learning	—	Learning Analytics: A Primer	—	International

Case Study: Learning Analytics at Cornell University

Course Overview

The *Learning Analytics* course at Cornell University has been offered in the Department of Information Science since 2018 by the first author. It enrolls around 200 students affiliated with six different colleges and over 20 different academic majors on campus. Students are mostly undergraduates in their final years (juniors, seniors) and

master's students in information or computer science, and a few doctoral students with an interest in education enroll each year. The course is designed to introduce students to various topics and methods in learning analytics and give them realistic opportunities to use education data to address practical issues and answer stakeholder questions. The course description summarizes the motivation and goals of the course:

> *Technology has transformed how people teach and learn today. It also offers unprecedented insight into the mechanics of learning by collecting detailed interaction and performance data, such as in online courses and learning management systems like Canvas. At the intersection of education and data science, learning analytics are used to make sense of these data and use them to improve teaching and learning. This course blends learning theories and methodologies covering a wide range of topics with weekly hands-on activities and group projects using real-world educational datasets. You will learn how learning works, major theories in the learning sciences, and data science methods. Students collect and analyze their own learning trace data as part of the course.*

Students are required to have foundational knowledge in programming and data analysis to enter the course because the course has a technical emphasis. However, the course does not assume any prior knowledge of educational or learning science theories. The official prerequisites state: This course is for undergraduate juniors, seniors, and graduate students interested in learning, education technology, educational data mining, and the broader implications of technology and data in education. Prior knowledge of probability and statistics (random variables, probability distributions, statistical tests, p values), data mining techniques (regression, clustering, prediction models), and fundamentals of programming is strongly recommended. Prior experience with the statistical programming language *R* is also recommended, as you will analyze data sets in *R* throughout this course.

The goal of the course is to prepare students for careers or further studies in education research, policy, and practice. By the end of the course, students are familiar with many foundational theories, contemporary trends, and widely used methods in the field of learning analytics and educational data mining. Moreover, they have gained experience working with raw, real-world datasets collected through education technologies, making informed decisions about how to clean the data, and interpreting the results of various methods that can be applied to the data to extract practical insights. Throughout the course, students consider the ethical, privacy, and equity implications of the applications they encounter to start forming a habit of considering these implications going forward. In line with these goals, the official learning objectives of the course are as follows:

- Explain key insights from learning science research and how learning works.
- Select and apply methods from educational data mining and learning analytics to analyze different kinds of educational data.
- Evaluate the results of different methods for different applications.
- Compare the strengths and weaknesses of methods for different applications.
- Identify potential benefits and risks of learning analytics for students, teachers, and institutions.

To accomplish these objectives, students complete readings, homework assignments, and group discussions on a weekly basis. The assignments are designed around authentic data extracted from educational technologies. For several assignments, students analyze data for their own class that is extracted from the course LMS. This makes the data and assigned questions to answer using the data especially personally relevant to students. The types of assignments are discussed in the next section and the strategy for incorporating learning analytics practice into the curriculum is discussed in the following section.

Course Structure

Students encounter a new topic in most weeks of the course. The lecture, readings, discussion section, and homework or group assignments during that week focus on the topic. What topics are included and how much time they receive represents a value judgment by the instructor. The topics can change over time as priorities shift and should be informed by an understanding of students' prior knowledge coming into the course and their career goals. The following topics are currently covered in the course: overview of what learning analytics is and why it matters; ethical and privacy considerations; how learning works; causal inference and A/B testing; multimedia learning and video analytics; assessments, psychometrics, and knowledge tracing; supervised and unsupervised predictive models; self-regulated learning; emotional learning analytics; learning analytics dashboards; and curriculum analytics.

In a typical week, students participate in the lecture which motivates the topic and the assignments for the week. They complete the readings and answer reading comprehension questions in the LMS to check their understanding, then they post a written summary on the discussion board and respond to another student's summary. The reading reflection posts and comments encourage students to identify and explain the core ideas from the week's readings, and compare their ideas to those of other students in the course. The specific reflection prompt in most weeks is "What are 3 things that you learned from the readings that you would tell someone who has not read them? Comment on someone else's reflection post to highlight an interesting takeaway that you had not previously thought of." Eager students who complete the reading reflection early tend to post longer and more thoughtful reflections, which are immediately visible to all other students and thereby establish a social norm to reflect deeply. Students participate weekly in small-group discussion sections led by a teaching assistant to talk about the readings and homework assignment. The homework assignment for the week is either an individual or team mini-project that involves data analysis in most weeks. There are three mini-projects in this course that require students to work as a team and coordinate to solve a problem. Teams are formed at the beginning of the course and they persist for the duration of the course. This ensures that every student has a close group of peers who they can ask for help even if they are from an underrepresented major in the course. Persistent teams give students an opportunity to develop a group culture and collective intelligence to tackle more challenging mini-projects later in the course. Students are assigned into groups of five based on their chosen discussion section (students enroll in one of many sections to fit their schedule) and responses to questions on the required start-of-course survey. Teams are assigned within sections to especially balance prior experience with the R statistical programming language, such that all teams have a similar average level of prior experience.

The course follows a mastery learning approach with explicit learning goals for each week and many opportunities for feedback. Students' final grades aggregate lecture and section attendance (10%), reading comprehension checks (10%), reading reflection posts/comments (10%), homework assignments (55%), and group projects (15%). The first three components are intended to be formative and therefore given just enough weight for students to complete them. They are merely graded for completion to encourage continuous engagement with the course each week. Homework solutions are released 48 hours after the due date. There are no midterm or final exams. The key to success in the course is to keep up with the material each week and ask for help early. Students can get help during weekly office hours and discussion sections, through the discussion forum, and from their peers.

Course Content

Overview of Learning Analytics. This introductory week introduces students to the field of learning analytics and educational data mining and gets them set up with the *R* programming environment that will be used for the homework assignments. Students watch a video from SoLAR (https://youtu.be/OOZhMjneMfo) and two introductory articles on big data in education (Baker and Inventado, 2014; Fischer *et al.*, 2020). As a self-assessed homework, students load a dataset into *R* and generate a report with basic descriptive statistics using starter code posted online, including exploring the dataset and answering basic questions about it. The stated homework learning objectives are (1) Identify a dataset file format and use the appropriate function to load it, (2) Explore fundamental

properties of a dataset using basic functions in R, (3) Compute and visualize relationships between variables using correlations, histograms, boxplots, and scatterplots, and (4) Calculate and visualize student- and question-level quantities and relationships.

Ethics and Privacy. The week on ethics and privacy engages students with questions about what data in education is collected by whom for what purpose, how the data is used, and what biases could emerge in the process. Students watch Neil Selwyn's keynote address at LAK 2018 (https://youtu.be/rsUx19_Vf0Q), followed by his article on *Re-imagining 'Learning Analytics'* (Selwyn, 2020). Students also read two complementary overview articles on algorithmic bias and fairness (Baker and Hawn, 2021; Kizilcec and Lee, 2022). The readings are discussed in sections and raise important questions for students, which the course returns to regularly. There is no homework assignment to provide extra time to get familiar with *R* and start reading for next week's group project.

How Learning Works. The week provides students an introduction to how learning works, based on learning science research, following a popular book on the topic (Ambrose *et al.*, 2010). Most students in the course have never taken an education course, thought systematically about how they learn and how learning works, let alone principles for effective teaching. All students read the introduction chapter and then, as their first group assignment, they create a 10-minute recorded presentation as a team about one of the seven principles of how learning works covered in the book. Students upload and share their presentations with other students in the course and everyone watches one presentation for each of the seven principles. For this week's reading reflection, students post (and comment on) two concrete ways that they could apply principles in a gateway STEM course.

Causal Inference and A/B Testing. This week focuses on the value and process of causal inference using randomized experiments, or A/B testing, in education. Students learn about different ways of conducting random assignment and how to analyze data collected from a randomized experiment. Students read the first chapter from the *Book of Why* (Pearl and Mackenzie, 2020) and a review chapter of experiments in online courses (Kizilcec and Brooks, 2017). It is revealed that the prior week's materials had an experiment embedded where students either watched a TED talk about grit or read the transcript before answering the same set of ungraded questions about the talk. Deidentified data collected from this experiment is provided to students for their homework assignment. Students also learn how to create and analyze A/B tests. The stated homework learning objectives are: (1) Understand the difference between simple, complete, and block random assignment, and know how to implement them, (2) Check the balance of an experiment, (3) Analyze experimental data using a t-test, linear regression, and Wilcox test, and (4) Report results of an experiment.

Multimedia Learning and Video Analytics. The week covers multimedia learning theory, a cognitive theory of how people learn with different content and how content should therefore be designed, and video analytics, a method for analyzing video activity data to gain actionable insights about learning and teaching. Students read chapters from *e-Learning and the Science of Instruction* (Clark and Mayer, 2011, chaps 2, 4), a handbook chapter on video analytics (Mirriahi and Vigentini, 2017), and a seminal video analytics paper (Guo, Kim and Rubin, 2014). For the homework assignment, students analyze video analytics data from a MOOC lecture video, identify activity spikes and other notable watching patterns, interpret them by examining these event times in the video, and provide recommendations to the instructor for how the lecture video might be improved. The stated homework learning objectives are: (1) Explore the structure of video interaction data, (2) Identify parts of the video with increased activity, and (3) Decide what video analytics to report back to learners and instructors.

Assessments, Psychometrics, and Knowledge Tracing. The week covers knowledge and skills assessment, with an introduction to standardized test development and validation using psychometric methods (classical test theory, item response theory [IRT]), and Bayesian knowledge tracing (BKT). Students read a handbook chapter on measurement (Bergner, 2017) and an article about using IRT to analyze the force concept inventory (FCI), a widely used assessment in introductory physics classes (Wang and Bao, 2010). They watch an expert interview about BKT

with Neil Heffernan, and optionally read a related (Pardos and Heffernan, 2010). For the homework assignment, students evaluate the psychometric properties of a standardized assessment, the FCI, that all students completed in the start-of-course survey. The stated homework learning objectives are: (1) Score and prepare an assessment for psychometric analysis, (2) Evaluate basic psychometric properties of an assessment like difficulty and reliability, (3) Apply and interpret an exploratory factor analysis, (4) Fit a Rash model and interpret Item Characteristics Curves.

Predictive Modeling: Supervised. The week on supervised predictive modeling covers a variety of uses and methods for predicting learner behavior and learning outcomes, with a focus on early warning systems. Students learn about different types of models to choose from depending on the prediction task and available data. Students read a commentary about not forgetting that learning analytics is about learning (Gašević, Dawson and Siemens, 2015), a handbook chapter on predictive modeling (Brooks and Thompson, 2017), and watch a short talk about the bias-variance tradeoff in educational data science (Doroudi, 2020). The homework assignment is to engineer features from a math tutoring dataset (ASSISTments Math 2004-05 downloaded from DataShop; https://pslcdatashop.web.cmu.edu/) and fit several simple predictive models (linear/logistic regression, kNN, Naive Bayes, regression/classification trees, and random forest) to predict student dropout and the number of questions they eventually complete. Students compare model performance, iterate on features, and interpret their findings. The stated homework learning objectives are: (1) Understand how to identify a problem that can be encoded as a prediction task, (2) Identify appropriate outcome variables and predictor variables, (3) Create new features based on existing data, and (4) Build and evaluate several different prediction models. The homework prepares students for the predictive modeling group assignment that is due the following week. In teams, students build an early alert model for students in this course using de-identified LMS data (raw clickstream, assignment-level grades) collected up to this point. The goal is to predict who does not submit the most recent homework on time 24 hours before the deadline. Students engineer features for different time periods to predict missed submissions each week only using data up to 24 hours before that week's deadline. They compare different modeling approaches and choose the best performing one, incentivized by extra credit for the two teams with the highest f1 score. The team writes a reflection on their experience and reasons they would (not) recommend using the model in class.

Predictive Modeling: Unsupervised. The week on unsupervised predictive modeling has students learn about finding patterns in data using methods such as cluster analysis and dimensionality reduction, and how they are used for understanding how learning behaviors and performance differ across groups of students. Students watch video explanations of k-means and hierarchical clustering before reading two articles about clustering learners in MOOCs (Ferguson and Clow, 2015; Khalil and Ebner, 2017). As the predictive modeling group assignment is due this week, students only receive an ungraded activity that guides them through performing dimensionality reduction with principal component analysis (PCA) and k-means clustering. They use student activity data from the same ASSISTments dataset to find groups with similar engagement and performance in five steps: (1) Roll up the data to student-level variables to cluster, (2) Check correlations and reduce the dimensionality of the dataset with PCA, (4) Apply k-means clustering for different values of k, (5) Interpret the findings.

Self-regulated Learning. The week covers self-regulated learning (SRL) theory, measurement, and interventions. Students learn about SRL phases and strategies, the use of self-report compared to clickstream data to detect SRL, and specific interventions focused on strategic plan-making and resource use. Students read a handbook chapter on learning analytics for SRL (Winne, 2017) and an article on strategic resource use interventions (Chen *et al.*, 2017), and watch a recorded interview with the study's lead author. The homework assignment has students search for evidence of established SRL strategies in the course's behavioral data and connect it to students' self-reported SRL strategies on the start-of-course survey (Kizilcec, Pérez-Sanagustín and Maldonado, 2017). Students propose ideas for features for each strategy, engineer them using the clickstream data, and examine how well they predict self-reported SRL strategies. This prompts students to realize the importance of instrumenting platforms to intentionally collect data about behaviors and processes like SRL. The stated homework learning objectives are: (1) Exploring response distributions of survey data, (2) Merging survey with behavioral data, (3) Engineering features that could

represent SRL strategies, (4) Checking if any behavioral features predict survey responses using a linear model. Students also keep a diary of their own SRL activities for one of their classes to raise their SRL awareness.

Emotional Learning Analytics. The week focuses on emotions in learning, ways of measuring learner affect, and applications that use affect data to support teaching and learning. Students watch Sidney D'Mello's keynote address at LAK 2017 about multimodal analytics (https://youtu.be/3sZmWyhK690) and read his handbook chapter on emotional learning analytics (D'Mello, 2017), an article about clickstream-based affect detection (Baker *et al.*, 2012), and an article about gaze-based detection of mind wandering (Hutt *et al.*, 2017). The homework assignment has students build a boredom detector using another ASSISTments dataset with validated affect labels (downloaded from https://sites.google.com/site/assistmentsdata/). The state homework learning objectives are: (1) Engineer features that can detect affect in a dataset, (2) Train a Random Forest model to identify boredom and plot the model's ROC curve, and (3) Make recommendations to teachers based on the features that are important.

Learning Analytics Dashboards. The week covers ways of communicating learning analytics to different stakeholders, such as students and instructors, with visualizations and summary statistics using a dashboard. Students learn about characteristics of an effective dashboard and how to develop one from needfinding to prototyping to implementation. Students read a handbook chapter on learning analytics dashboards (Klerkx, Verbert and Duval, 2017) and articles about student-facing (Bodily *et al.*, 2018) and teacher-facing dashboards (Echeverria *et al.*, 2018). Students also watch a tutorial video for R Shiny (https://shiny.rstudio.com/) and ggplot2 (Wickham, 2016), which they use for their final group assignment: creating a student or an instructor dashboard for a Cornell course that has provided deidentified clicker data combined with student grades. Student teams have two weeks to plan what information would be valuable to present and how to present it, draw mock-ups and get feedback, implement the data processing, visualizations, and dashboard using R Shiny, and write a report reflecting on their design choices. The stated homework learning objectives are: (1) Understand the structure of clicker data, (2) Create multiple different visualizations, (3) Design and implement an instructor or student dashboard, and (4) Critically evaluate your own dashboard design.

Curriculum Analytics and Academic Pathways. The final week focuses on curriculum analytics and academic pathways in the context of undergraduate programs. Students learn about higher education data for course and major choices, course search, grades, and other attributes, and how they can be used to inform students, instructors, advising staff, and academic leaders. Students read an article about measuring and interpreting undergraduate course consideration patterns (Chaturapruek *et al.*, 2021), watch a talk on facilitating course articulation for transfer students by Zachary Pardos (Pardos, Chau and Zhao, 2019), and on creating a lifelong learning marketplace by Mitchell Stevens (https://youtu.be/ehPs8qDs1V0). For the homework assignment, students analyze fully deidentified course enrollment records with grades. The stated homework learning objectives are: (1) Understand how course enrollment data is structured, (2) Identify hard course pairings using enrollment data, and (3) Identify course-major relationships to give students feedback about path dependencies.

Tools and Resources Used

The course uses edX Edge as the LMS because this makes it relatively easy to extract and provide LMS data to students. The edX data schema is less complex and requires less pre-processing to be usable by students compared to Canvas. A number of students have said that they appreciated the opportunity to try out a different LMS in this course: it gave them a better understanding of the nature of an LMS. edX Edge also facilitates the implementation of A/B testing and passing a hashed student identifier to a survey via the URL to conveniently connect survey responses to the behavioral data. Instead of the edX discussion board, the course uses Slack for general course updates and reminders, posting weekly reading responses and comments, and help-seeking in an asynchronous office hours channel. Each student team also creates a private channel for communicating amongst themselves. Students use direct messaging on Slack instead of email for any private or sensitive inquiries. To help students keep

track of the weekly lecture, section, content release dates, and deadlines, a calendar file is created that includes all of these events and exported as a .ics file, which students can easily upload into their preferred calendar application.

The course readings and video presentations are either publicly available (e.g., SoLAR Handbook of Learning Analytics; https://www.solaresearch.org/publications/hla-17/) or accessible through institutional networks. The course uses the statistical programming language *R* with the graphical user interface RStudio (https://www.rstudio.com/), though it could also be taught in Python which the students are more familiar with, but many of them appreciate the opportunity to improve their ability to use *R* in this course. Most of the educational datasets are either publicly available, such as the ASSISTments dataset, or collected from students in the course using the edX LMS and the start-of-course survey. Other datasets are obtained from other courses or institutions, such as the in-class clicker or video analytics datasets. The number of publicly available datasets is increasing thanks to public competitions with educational datasets and efforts to promote open science practices that include releasing de-identified data (e.g., ASSISTments data repository https://sites.google.com/site/assistmentsdata/ and CMU DataShop https://pslcdatashop.web.cmu.edu/).

Incorporating Learning Analytics Practice into the Course

Students benefit from the experience of working with authentic education data to answer personally relevant questions. In the words of one student in the end-of-course survey, "it felt like I was studying at the forefront of an emerging field and had a unique opportunity to participate and experiment with different ideas. I liked the freedom given to find solutions to problems." For many students, it is the first time that they think systematically about learning and teaching, the affordances of technology in this domain, and the opportunities and concerns that learning analytics bring. One student commented on this eye-opening experience, "I was introduced to a topic I had never even heard about. The psychological concepts presented throughout the course made me more aware of my own learning. It was also interesting to learn about the metrics (which I would have never thought of) that are used to assess and improve learning." Students who are interested in pursuing a career in data science are generally aware of opportunities in technology, financial services, medical, and marketing companies, but many of them are unaware of the options they have at education companies, as noted by another student: "The content of the course was really inspiring and made me think of data science in a completely different way. It inspired me to pursue a career and grad school education in learning analytics." A course on learning analytics can have a lasting impact on people's lives and lifelong learning practices by engaging them meta-cognitively with the process of learning and letting them discover what educational data is capable of and what its limitations are.

A recurring theme in the course is the cross-cutting consideration of ethics, equity, and culture. The weekly lecture and discussion highlight implications for student privacy, informed consent, data ownership, unintended consequences of well-intentioned interventions, questionable uses of student data, and randomized experiments in educational contexts to encourage students to think critically about how learning analytics affect people, institutions, and society. Students learn about current inequities in the education system and are encouraged throughout the course to attend to the ways that learning analytics applications might improve, perpetuate, or exacerbate them. Aside from algorithmic fairness considerations, the course lectures cover psychological theories (e.g., social identity and belonging, identity-based motivation, social norms, cognitive biases) that can help students understand how the users of learning analytics applications—students, teachers, staff, administrators—may act in ways that are not inclusive and potentially reinforce inequities. Finally, it is important for a course on learning analytics to acknowledge and reflect the diversity of cultural perspectives and practices for learning and teaching around the world. While the course content is US-centric, the lectures highlight examples from other cultures, and students who come from around the world are encouraged to share their educational experiences and contextualize the course content within their cultural frame of reference.

The Future of Learning Analytics Education

Learning analytics education today is highly distributed—geographically, methodologically, and across disciplines. This is a promising indicator of the growing popularity and strong value proposition of learning analytics to a variety of stakeholders beyond academia. It is important that the community maintains a balance of upholding its core principles while simultaneously expanding to accommodate and reap the benefits of a growing list of partner disciplines. Success in this regard will largely stem from the manner with which future generations of learning analytics researchers and practitioners are trained. We are at an important crossroads now as we come to terms with this fact: *none of the current leaders in the learning analytics field were trained as learning analysts*. The term "learning analytics" simply did not exist, and neither did learning analytics curricula. Each leader was drawn to the unprecedented troves of educational data made possible by the advent of large-scale open online learning, carrying with them their disciplinary practices along with shared passions and curiosities for the science of learning. While it only takes a small group of visionaries to invent a discipline, it takes a highly coordinated community to grow and nurture one.

From reviewing and comparing the programs identified in our survey of the landscape and closely examining the curriculum of a specific course, we distilled the following takeaways and recommendations for the community to reference when designing and building learning analytics courses. We do not intend to build walls around rigid guidelines defining the discipline, but rather to encourage current and future educators and learners to consider promising approaches and innovations in this domain.

- **Build a theoretical foundation** - Before students are asked to conduct any analyses or learn a new programming language for data processing, it is critical that they first develop a strong foundational understanding of the field from its inception to the state of the art. This enables students to properly justify and contextualize their own analyses by intentionally selecting the types of problems, questions, and methods they engage with. It encourages students to be critical consumers and informed producers of learning analytics insights. Developing this theoretical foundation will help students achieve a new literacy for peer-reviewed quantitative research articles that can allow them to stay up-to-date in a fast-moving field. Which specific theories should be taught as part of this foundation varies across programs for now but may converge into a theoretical core in the future.

- **Include practical quantitative elements** - Courses should be designed such that students have at least some hands-on experience with educational data. Ideally, this would also entail using a programming language such as R or Python (instructors should select whichever language fits the context of the program), but spreadsheet-based tools like Microsoft Excel are also useful for conveying the same core ideas. Students benefit from learning and honing the skills of using a programming language to conduct their own analyses, however simple they may be because they learn about all of the decisions that go into any such analysis. This provides them with the awareness, literacy, and understanding to evaluate learning analytics findings and the methods used to arrive at them.

- **Make learning analytics self-relevant** - Learning analytics courses can draw students from a wide variety of backgrounds, as reflected in the diversity of departments that offer them. Instructors should embrace this diversity by encouraging students to bring their own interests and experience to the table. For example, when students work on analytical (research) projects during a course, it can be a meaningful experience if they have the option to bring their own dataset—whether it is one from their primary job, volunteering, or one found online—or the possibility to analyze their own individual and classroom-level data, as illustrated in the case study.

- **Encourage critical reflection** - Learning analytics courses may be the first time students learn about all the data generated based on their interactions and performance, how these data might be used in practice, and potential randomized experiments embedded in their courses. This can lead students to raise concerns over privacy, ethics, and regulations. These concerns should not only be addressed but welcomed and openly discussed in the course. For learning analytics to continue developing as a field, instructors and researchers need to have an ear to the ground and understand students' concerns and how they make sense of them. Not only will it help address the concerns, but it can also inform future research and product development.

- **Open-source course materials** - To advance our collective understanding of learning analytics education, we encourage instructors to make their syllabi and resources available online whenever possible. Not only does this increase the reach and accessibility of learning analytics materials to broader audiences, but it also fosters a sense of community among instructors who can learn from and build off of one another's teaching approaches. Instructors can further participate in opportunities to exchange ideas and materials about teaching learning analytics at conferences or other social convenings.

A relatively nascent field, learning analytics benefits from the flexibility to respond to emerging issues in education and digital technology. When ethical concerns around big data and technology firms gained traction in the public sphere, the learning analytics community swiftly began devising frameworks and publishing research about the role of ethics and privacy in the collection and use of educational data (Slade and Prinsloo, 2013). This has also encouraged efforts to prioritize teaching a code of ethics in learning analytics courses (Prinsloo and Slade, 2017). Moreover, in response to the 2020 racial justice movement to address systemic issues of justice, equity, diversity, and inclusion, the learning analytics community committed to "identify and eliminate racial disparities…[and] mobilise our expertise and connections with communities to actively contribute to the hard work of promoting social justice and dismantling injustices in education" (https://www.solaresearch.org/2020/06/statement-of-support-and-call-for-action/). Learning analytics educators can contribute to this cause by exploring ways to teach learning analytics to empower students to eliminate disparities and promote social justice. Finally, learning analytics applications are increasingly adopted in parts of the world with cultural norms and values about teaching and learning that differ from those in Western nations, including differences in epistemological beliefs, pedagogical orientation, uncertainty tolerance, and methods for consensus building (Kizilcec and Cohen, 2017; Baker *et al.*, 2019; Rizvi *et al.*, 2022). Including different cultural perspectives in the learning analytics community is essential to building an inclusive body of knowledge and avoiding imposing Western educational values in other contexts. Educators have the power and arguably a responsibility to show this diversity of thought and practice to their students by not only selecting readings, case studies, datasets, and class projects that are culturally relevant but also ones that expose them to unfamiliar cultures. Given the preponderance of learning analytics education programs from Western countries we observe in our review, there is a need to intentionally check that our community is promoting teaching and learning of learning analytics in a culturally inclusive manner. We hope that the insights and guidance provided in this chapter can facilitate the development of new educational programs around the globe.

References

Ambrose, S.A. *et al.* (2010) *How Learning Works: Seven Research-Based Principles for Smart Teaching*. John Wiley & Sons.

Baker, R.S. *et al.* (2012) 'Towards Sensor-Free Affect Detection in Cognitive Tutor Algebra', in *Proceedings of the International Conference on Educational Data Mining (EDM)*. *International Conference on Educational Data Mining (EDM)*, International Educational Data Mining Society. Available at: https://books.google.com/books/about/Towards_Sensor_Free_Affect_Detection_in.html?hl=&id=tw4JvwEACAAJ.

Baker, R.S. *et al.* (2019) 'Culture in computer-Based Learning systems: Challenges and opportunities', *Computer-*

Based Learning in Context, 1(1), pp. 1–13.

Baker, R.S. and Hawn, A. (2021) 'Algorithmic Bias in Education', *International Journal of Artificial Intelligence in Education*, pp. 1–41.

Baker, R.S. and Inventado, P.S. (2014) 'Educational data mining and learning analytics', in *Learning Analytics*. New York, NY: Springer New York, pp. 61–75.

Bergner, Y. (2017) 'Measurement and its Uses in Learning Analytics', in *Handbook of Learning Analytics*. Society for Learning Analytics Research (SoLAR), pp. 35–48.

Bodily, R. *et al.* (2018) 'The design, development, and implementation of student-facing learning analytics dashboards', *Journal of Computing in Higher Education*, 30(3), pp. 572–598.

Brooks, C. and Thompson, C. (2017) 'Predictive modelling in teaching and learning', in *Handbook of Learning Analytics*. Society for Learning Analytics Research (SoLAR), pp. 61–68.

Chaturapruek, S. *et al.* (2021) 'Studying undergraduate course consideration at scale', *AERA open*, 7, p. 233285842199114.

Chen, P. *et al.* (2017) 'Strategic Resource Use for Learning: A Self-Administered Intervention That Guides Self-Reflection on Effective Resource Use Enhances Academic Performance', *Psychological Science*, 28(6), pp. 774–785.

Clark, R.C. and Mayer, R.E. (2011) *e-Learning and the Science of Instruction: Proven Guidelines for Consumers and Designers of Multimedia Learning*. John Wiley & Sons.

Clow, D. (2013) 'An overview of learning analytics', *Teaching in higher education*, 18(6), pp. 683–695.

D'Mello, S.K. (2017) 'Emotional Learning Analytics', in *Handbook of Learning Analytics*. Society for Learning Analytics Research (SoLAR), pp. 115–127.

Doroudi, S. (2020) 'The bias-variance tradeoff: How data science can inform educational debates', *AERA open*, 6(4), p. 233285842097720.

Echeverria, V. *et al.* (2018) 'Exploratory versus explanatory visual learning analytics: Driving teachers' attention through Educational Data Storytelling', *Journal of Learning Analytics*, 5(3), pp. 72–97.

Ferguson, R. and Clow, D. (2015) 'Examining engagement', in *Proceedings of the Fifth International Conference on Learning Analytics And Knowledge. LAK '15: the 5th International Learning Analytics and Knowledge Conference*, New York, NY, USA: ACM. doi:10.1145/2723576.2723606.

Fischer, C. *et al.* (2020) 'Mining Big Data in Education: Affordances and Challenges', *Review of Research in Education*, 44(1), pp. 130–160.

Gašević, D., Dawson, S. and Siemens, G. (2015) 'Let's not forget: Learning analytics are about learning', *TechTrends : for leaders in education & training*, 59(1), pp. 64–71.

Guo, P.J., Kim, J. and Rubin, R. (2014) 'How video production affects student engagement', in *Proceedings of the first ACM conference on Learning @ scale conference. L@S 2014: First (2014) ACM Conference on Learning @ Scale*, New York, NY, USA: ACM. doi:10.1145/2556325.2566239.

Hutt, S. *et al.* (2017) 'Gaze-Based Detection of Mind Wandering during Lecture Viewing', in *Proceedings of the International Conference on Educational Data Mining (EDM). International Conference on Educational Data Mining (EDM)*, International Educational Data Mining Society.

Khalil, M. and Ebner, M. (2017) 'Clustering patterns of engagement in Massive Open Online Courses (MOOCs): the use of learning analytics to reveal student categories', *Journal of Computing in Higher Education*, 29(1), pp.

114–132.

Kizilcec, R.F. and Brooks, C. (2017) 'Diverse big data and randomized field experiments in MOOCs', in *Handbook of Learning Analytics*. Society for Learning Analytics Research (SoLAR), pp. 211–222.

Kizilcec, R.F. and Cohen, G.L. (2017) 'Eight-minute self-regulation intervention raises educational attainment at scale in individualist but not collectivist cultures', *Proceedings of the National Academy of Sciences of the United States of America*, 114(17), pp. 4348–4353.

Kizilcec, R.F. and Lee, H. (2022) 'Algorithmic Fairness in Education', in Wayne Holmes, K.P.-P. (ed.) *The Ethics of Artificial Intelligence in Education: Practices, Challenges, and Debates*. Routledge.

Kizilcec, R.F., Pérez-Sanagustín, M. and Maldonado, J.J. (2017) 'Self-regulated learning strategies predict learner behavior and goal attainment in Massive Open Online Courses', *Computers & Education*, 104, pp. 18–33.

Klerkx, J., Verbert, K. and Duval, E. (2017) 'Learning Analytics Dashboards', in *Handbook of Learning Analytics*. Society for Learning Analytics Research (SoLAR), pp. 143–150.

Lazer, D. *et al.* (2009) 'Social science. Computational social science', *Science*, 323(5915), pp. 721–723.

Mirriahi, N. and Vigentini, L. (2017) 'Analytics of learner video use', in *Handbook of Learning Analytics*. Society for Learning Analytics Research (SoLAR), pp. 251–267.

Pardos, Z.A., Chau, H. and Zhao, H. (2019) 'Data-Assistive Course-to-Course Articulation Using Machine Translation', in *Proceedings of the Sixth (2019) ACM Conference on Learning @ Scale. L@S '19: Sixth (2019) ACM Conference on Learning @ Scale*, New York, NY, USA: ACM. doi:10.1145/3330430.3333622.

Pardos, Z.A. and Heffernan, N.T. (2010) 'Modeling individualization in a Bayesian networks implementation of knowledge tracing', in *User Modeling, Adaptation, and Personalization*. Berlin, Heidelberg: Springer Berlin Heidelberg (Lecture notes in computer science), pp. 255–266.

Pearl, J. and Mackenzie, D. (2020) *The Book of Why: The New Science of Cause and Effect*. Basic Books.

Prinsloo, P. and Slade, S. (2017) 'Building the Learning Analytics Curriculum: Should we Teach (a Code of) Ethics?', in Bergner, Y. et al. (eds) *CEUR Workshop Proceedings*. *Workshop on Building the Learning Analytics Curriculum*. Available at: http://ceur-ws.org/Vol-1915.

Rizvi, S. *et al.* (2022) 'Culturally adaptive learning design', in Rienties, B. et al. (eds) *One World Learning*. Routledge, pp. 103–115.

Sawyer, K.R. (2005) *The Cambridge Handbook of the Learning Sciences*. Cambridge University Press.

Selwyn, N. (2020) 'Re-imagining "Learning Analytics" … a case for starting again?', *The Internet and Higher Education*, p. 100745. doi:10.1016/j.iheduc.2020.100745.

Siemens, G. (2013) 'Learning analytics', *The American behavioral scientist*, 57(10), pp. 1380–1400.

Slade, S. and Prinsloo, P. (2013) 'Learning analytics: Ethical issues and dilemmas', *The American behavioral scientist* [Preprint]. Available at: https://journals.sagepub.com/doi/abs/10.1177/0002764213479366.

Tursunbayeva, A., Di Lauro, S. and Pagliari, C. (2018) 'People analytics—A scoping review of conceptual boundaries and value propositions', *International journal of information management*, 43, pp. 224–247.

Wang, J. and Bao, L. (2010) 'Analyzing force concept inventory with item response theory', *American Journal of Physics*, 78(10), pp. 1064–1070.

Wickham, H. (2016) *ggplot2: Elegant Graphics for Data Analysis*. Springer.

Winne, P.H. (2017) 'Learning analytics for self-regulated learning', in *Handbook of Learning Analytics*. Society for Learning Analytics Research (SoLAR), pp. 241–249.

Kizilcec, R. F., Piech, C., & Schneider, E. (2013). Deconstructing disengagement: analyzing learner subpopulations in massive open online courses. In Proceedings of the third international conference on learning analytics and knowledge (pp. 170-179).

Wagner, E., & Ice, P. (2012). Data changes everything: Delivering on the promise of learning analytics in higher education. Educause Review, 47(4), 32.