Exploring Learning Analytics:
The Role of Data Regulation and Privacy

Aishwarya V. Nadgauda

Thesis Advisor: Prof. Gad Allon
Engineering Advisor: Dr. Sampath Kannan

EAS 499: Senior Capstone Thesis
University of Pennsylvania
School of Engineering and Applied Science
Department of Computer and Information Science
May 1, 2019

Table of Contents

Introduction:

There isn't an industry today that isn't being turned on its head with the potential of data analytics. It seems that everything is on the brink of being discovered, and insights we can gain by analyzing the vast amounts of data that has been collected is all that's needed to put us over the edge. We live in a day and age where every single advertisement that we see in the little pop up banner on our Google Chrome screen is targeted and customized.

The roots of data analytics as we understand it today began in the late 1960s when we saw for the first time computers become a part of the decision-making process. This was happening as a result of the advent of as well as data warehouses. There were also considerable strides in software and hardware technology made during this period of time that enabled faster processing of data as well as cheaper storage of data.

At the heart of data analytics lies statistics. Data has been collected and sifted through to garner new insights for long before data analytics as we understand it now existed. Relational databases played a critical role in bringing together datasets. The expansion into non-relational databases has allowed for greater flexibility in connections and correlations that could be discovered within the data.

The 1980s brought with it the data warehouses which had significant implications for data storage and, as a result of that, for the scope of data analytics. This was a remarkable leap forward from data being stored on hard disk drives. Data warehouses allowed not only for more data storage but also for faster access to data. As a result of these advances, the next decade is when we see data analytics pick up.

Data analytics has permeated every industry and the education space is no exception. In recent years we have seen technology become commonplace in the classroom. Notebooks and textbooks are rapidly replaced by tablets, laptops, smartphones etc. We see the introduction of a new medium through which we interact with the student. This in turn allows for collection of more data about these students and how they are interacting with the material to be collected.

At the highest level we see that storage and computational capabilities have seen an increase. These capabilities have opened up the possibility of doing a lot more with the data than was previously imagined. The combination of this with the fact that there has been an increase in data collection in the classroom is that we see the emergence of educational data mining. The goal of educational data mining is to work towards improving the quality and accessibility of education.

As the potential for this to occur has increased, funding going into the educational data mining space is increasing as well. The Hewlett and Gates foundation allocated almost $25 million to higher education projects of which $3 million was focused on projects in the analytics space [1]. Even the government has begun to allocate more money into this space. It is seen as having the capability to positively impact the economy at a global level. During

Obama's presidency, the MyData initiative emerged which worked towards giving students digital copies of their academic record which they could then download [2].

The key consumers of education technology are either students or learners. Students are individuals in a traditional classroom setting who are interacting with education technology through an assessment on their tablet. Learners can be anyone consuming material from an education technology platform outside of a classroom setting. An example of a learner would be an individual who is working as a software engineer and uses a course online to learn a new coding language to help them with a project at work.

While educational data mining considers a broad landscape of ways in which technology and education can collaborate, we are honing in on learning analytics specifically. Learning analytics is a subset of educational data mining focused on gaining insights from the data to increase the chances of a student's success.

In educational data mining, traditional analysis (e.g. clustering) are performed on sets of data. Learning analytics is doing this, but also considering additional factors on top of this. It incorporates instruments such as visualization tools and social network analysis. The premise of learning analytics is going a step further than traditional education data mining. There is an aspect of it that involves trying out the theories and ideas that come about as a result of the data mining and seeing how effective they are.

Given the potential of learning analytics, we see it proliferated in multiple ways from students in kindergarten through executives in companies. Learning analytics is a multidisciplinary field drawing upon expertise from individuals across both sectors (education and technology). Successful ideas and products in this industry require sufficient knowledge about a range of topics from what the classroom setting looks like to the regulation around student privacy to how to build a website. About a decade ago, we see the space getting more attention at an international level. The first international conference on educational data mining convened in 2008. This was followed by the first learning analytics and knowledge conference being held in 2011 [3].

Scope of the Paper:

Over the course of this thesis, we will begin by understanding the most common technological approaches to learning analytics and where the data is being sourced from such that this is feasible. From there, we consider the current applications that have been derived from this technology in industry and the factors (such as stakeholders, business models, market forces and regulation) that are shaping the way in which companies in this space work. As the data that is utilized in analytics plays a very fundamental role in setting the path for the potential of the technology, the issue of data privacy is critical. Finally, we will consider how regulation in the space influences data privacy standards and the implications of that for the growth of learning analytics applications in the field. The argument will be made that regulation should allow for collaboration in the field of data analytics in order to create learning environments that benefit the most students.

Technology of Learning Analytics

The fundamental approaches to learning analytics mirror those in the larger data analytics space that span a multitude of industries.

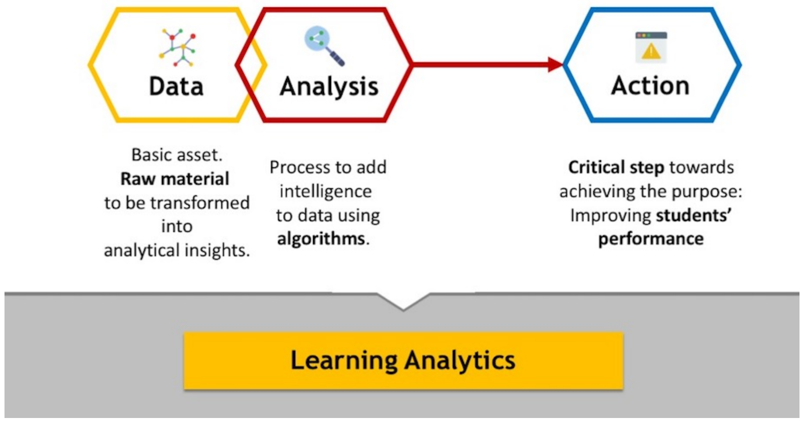


Figure 1: Learning Analytics
Source: IADLearning [4]

As can be seen in Figure 1, there are three critical components to analytics: data, analysis and then action. Data is referring to the base information which undergoes the process of analysis which allows there to be additional insights that one could not get just by looking at the data. Finally, there is the action component which is how change a process or react to the information that we have garnered. This step is critical to the success of learning analytics as it is only through clear action taken that students can get the benefits that come about as a result of the first two steps.

The approaches to the second step, analytics, that are used in learning analytics are classification, clustering and social network analysis. These methodologies serve as approached to prediction. Prediction is when there is a data point that we want to infer, and this is done by considering an aggregate of other data points that have been looked to. Predictive modeling helps to fill in the gaps that the data may have. To build out a prediction model, the most commonly used methodology or the analysis step is classification.

Classification:

Classification is essentially a predictive model where we are giving a label to each object in the data set that is being considered. Examples of this include decision trees, neural networks, naive Bayesian classification, support vector machines and k-nearest neighbor classification. The goal of this classification is to be able to take an unidentified object and put it in the correct grouping. As we build out these models, cross-validation plays an important role[3]. This is integral in deeming the accuracy of the model at each point in time. For example, in the financial industry this is what is used in loan groupings — classification allows one to be able to consider a loan application and identify it with the label low, medium or high risk.

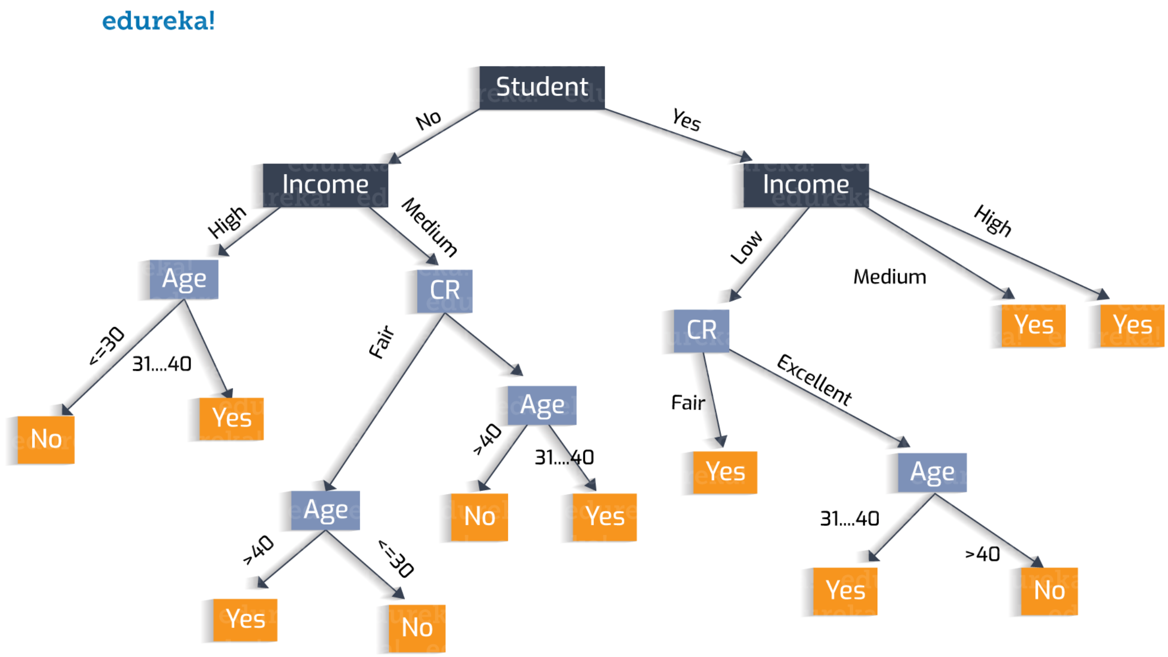


Figure 2: Sample Decision Tree on Students
Source: Edureka

A specific example of classification, a decision tree model, is very commonly used in the learning analytics space. The big picture idea of a decision tree is that there are decision nodes, and based on the outcome at that point, we move to one of two other decision nodes. Figure X displays an example of how a learner could be modeled through a decision tree. The factors that one would be considering in this case are income, age and credit rating.

The splits in the decision tree model are chosen based on what gives you the most significance at any point in time. So you are essentially prioritizing the splits that give you the most information to the top of the decision tree. This allows you to give different weightage to various attributes. The end result of a decision tree model is a series of very simple rules that can be used in classification.

Classification in Learning Analytics:

Looking at how we can take this model of classification and translate it into the learning analytics space, the first step is to understand what the labels of objects would be. Examples of this in the learning analytics world are line items such as standardized test scores or grades. The value of prediction can really be gained by being able to identify when it is most impactful to intervene in a student's learning process. Additionally, it can provide constructive insight into where a student's current understanding of the material lies. Predictive modeling through classification can further be used to gain a perspective on factors that contribute to a certain outcome. An example of this is looking at what determines whether a student graduates high school.

A subset of classification that is utilized frequently in the learning analytics sector is latent knowledge estimation. At the core of this approach is that we are trying to measure a student's knowledge, which is not directly measurable and must be inferred based on a student's performance [3]. Decisions that administrators and educators would make based on this information would include whether or not the individual should progress in the curriculum or if a teacher should intervene.

The two algorithms most commonly employed for latent knowledge estimation are Bayes Nets and Bayesian Knowledge Tracing. When the "knowledge" that is being considered is more complex, we turn to Bayes Nets. In Bayes Nets, the data is structured in an acyclic graph, where conditional dependencies are represented by edges. If it is a case where, by looking at one problem or one group of problems, we are trying to see if there is a single skill that the student has been able to master, then Bayesian Knowledge Tracing [5] is used. Bayesian Knowledge Tracing has four output parameters:

- *p-init → What is the probability of the student having an understanding of the skill prior to being introduced to this material?*
- *p-transit → What is the probability of the student showing that they know the skill given that they have the chance to apply it?*
- *p-slip → What is the probability of the student performing an error when they are applying the skill?*
- *p-guess → What is the probability of the student performing the skill currently by chance?*

Clustering:

The key difference between clustering and classification is that in classification one has an initial data set that you are starting with. This is called the training set which is comprised of objects that are already labeled into groups. With clustering, these labels do not exist in advance. Here you go through the process of grouping objects together such that you are minimizing the "distance" between them. The distance being considered here is some measure of the differences between the groupings.

In clustering, there are different levels and sizes at which the clustering can occur—students, grades, schools student behaviors etc. There are two key approaches to clustering: hierarchical agglomerative clustering (HAC) and non-hierarchical. The difference between the two approaches lies in that the HAC model operates under the assumption that clusters can cluster together, which non-hierarchical views clusters as entirely separate entities that cannot be grouped together further[3].

Clustering in Learning Analytics:

With clustering, we have the potential to build a model that is able to reproduce the way in which students are interacting with online course material. This is a necessary step for the technology to then be able to predict how a certain student profile will respond to material delivered in a specific manner[15].

With clustering, one of the most complex steps is determining the number of groups or categories into which the data should be bucketed. The attributes upon which one chooses to group must be such that there is no overlap amongst the buckets. In measuring how "good" a given clustering is, we calculate the distance between the groups. Essentially what is being maximized is homogeneity within a group and heterogeneity between groups.

Social Network Analysis:

In a social network analysis, you are looking at the information that you are taking away from learning analytics as a graph. This tends to be used more often in a quantitative analysis. This can be helpful as in the learning analytics field, as is true across analytics in multiple industries, we often run into the issue of "data right, but information poor"[6]. It is difficult to go from raw data to helpful insights that can actually guide the decisions that are being made.

This is particularly helpful in a subset of learning analytics called social learning analytics. Social learning analytics is centered around the idea that the learning environment is not a siloed experience between a student and instructor. Research in this space looks into how social dimensions involving one's network influence the process of learning [20]. Social network analysis is a critical component of being able to recognize the networks that exist and how they interact with each other.

Supplemental Technologies:

Learning analytics involves taking the insights gained from data analysis methods, such as clustering and classification, and considering how they can be fed into an application that allows its insights to positively impact students and learners.

One way through which this is done is information visualization where we take statistical information (such as how many times a page was visited or what percentage of material an individual has looked at) and present it in a way that is easier for the individual who is processing the information to understand what it is they are looking at. From this information, we can more effortlessly communicate the answers to questions such as "what are the key takeaways?". It essentially draws attention to the relevant information such as highs and lows.

An additional way in which learning analytics utilizes the awareness brought about through such platforms is by serving as the foundation for a recommender system. The two main types of recommender systems are content based recommendations and collaborative filtering [6]. Content based recommendations is based off of what the user has already been exposed to. Collaborative filtering is based on what other users who have viewed similar things to what the current user has viewed. Optimally, a combination of the two recommendation systems are used.

The recommendation systems serve as the foundation for personalized adaptive learning. Adaptivity is when course material can be changed, but it is based on predetermined guidelines. Adaptability is the personalization of course material, and this is where learning analytics has the potential to grow a lot [6]. There has been a lot of focus in the learning analytics space around this as is evident through examples such as Intelligent Tutoring System and Adaptive Hypermedia Systems.

As technological advancements have taken place in personalized adaptive learning, we have seen the emergence of the Personal Learning Environment, which is focused on the idea of adaptability. The main idea here is that the learner should be the one guiding the environment in which they are learning [6]. As such, they are not restricted by the environment or instructor in which they are learning. Learning analytics is critical for this to happen. This is how we have the potential to give material to the learner in an order and manner in which they are most likely to be successful in understanding and retaining the information presented to them. An important part of this is being able to tell where the learner is currently in their understanding of certain material.

Data Collection and Analysis:

To varying degrees, data has been used to drive learning and interaction in the classroom prior to the field of learning analytics picking up speed. Most examples of this occur at a small scale, such as considering a server log to see which test questions students had the most trouble with. Only utilizing data at this level of granularity fails to reap the benefits of a more global and systematic approach to learning analytics.

As mentioned before, the data sources that are pooling into learning analytics play a critical role in driving the potential for growth. The quantity and quality of the data is the baseline for the caliber of the insights that are derived.

Sources of data include centralized education systems and distributed learning environments. Centralized education systems include learning management systems such as Blackboard which serves as an online platform to supplement learning. Through this informational logs of student activities can be collected. Questions such as “What material are they accessing?” or “How have they done on tests?” can be answered. A distributed learning environment is one in which the instructor and students can be in different location and the curriculum is still delivered.

The data utilized in analytics primarily come from student information systems and learning management systems. Student information systems such as PowerSchool offer a way for schools to monitor not only student performance but also offers additional features such as scheduling[21]. Learning management systems such as LearnUpon serve as a platform for their customers to create and deliver content to their students, whether those be executive or school children.

Through a student information system, information can be collected on the student’s demographics, grades, socioeconomic standing etc. Additionally, through an integrated library system, we can get data points around the resources an individual is using. Electronic textbooks allows for potential data collection around the way in which a student is consuming certain material available to them. Fundamentally, the influx of technology in a classroom setting is driving the feasibility of this data collection and altering the way in which we are approaching delivering information to students.

To further our capacity to build out personalization in this space, more diverse data sets must be built out. One potential sources of data involving interaction in a physical learning space would be through wearable devices that collect data. As it is actually used right now, majority of learning analytics relies on data that is automatically captured. When we consider projects in the analytics space that are of much larger magnitude in industries such as healthcare and retail, where learning analytics aspires to to be, there is considerable amount of manual effort in terms of data collection that goes into it. Particularly with healthcare, a significant portion of the data around a human's general health is collected manually by a doctor. This data is then conglomerated to study. As the applications of the space expand, the models would ideally account for a multitude of data inputs, including both automated and human sources.

This point gains particular importance as we turn to knowledge domain modeling. A domain model is "defining causal relationships within data, based on a deep understanding of a domain, allowing us to anticipate consequences and causes of actions and to learn new causal relationships hidden in data" [22]. With knowledge domain modeling, we hone in on the personalization aspects of the applications in learning analytics[14]. In discussing knowledge domain modeling, the first step is recognizing that a multidisciplinary approach is crucial to achieving success in this regard. For example, curriculum developers should drawn upon this to capture what content each type of student would be most inclined to benefit from.

Current Limitations of Data Collection:

As mentioned above, the main sources of data currently come from learning management systems and student information systems. From this data, the conclusions that are often developed are as follows: the student who logs into a learning management system more frequently is more likely to have a higher score on the assessments presented in class. While there is value to this insight in itself, if we view the goal of learning analytics as studying data in order to create a learning environment that is more beneficial, these types of insights are not conducive to that. Simply telling a lower performing student to log into the learning management system more frequently would not aid that student in becoming more successful. In order to move towards more actionable insights, the sources from which we collect data must be expanded to include more information on how the student is actually learning.

Current Applications:

Success in the field of learning analytics requires a multi-disciplinary approach. The strides made from a technological perspective are just one of the prongs of further development in learning analytics. The landscape of the industry plays a critical role in dictating how rapidly growth can occur and where in the pain points lie. We will look at where the field currently stands, consider who the the core stakeholders in learning analytics are and how their goals intersect and where they are at odds. Then considering the main sects of applications actually in industry, we will examine how companies have leveraged these applications to generate revenue and where there is potential for future growth.

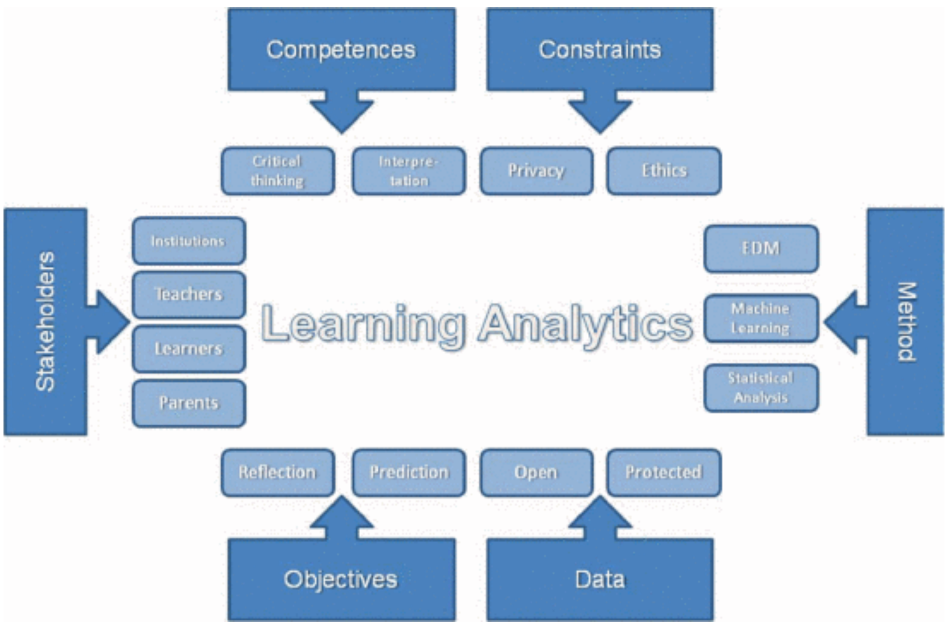


Figure 3: Proposed Learning Analytics Framework
Source: IEEE [23]

Current State:

Learning analytics encompasses a range of techniques which drive applications that alter the way in which our learning environment is informed. As it currently stands, the learning analytics space is not meeting its potential in terms of insights that it can offer its consumers. It is evident through a lack of operational efficiency that the educational sector is falling behind other industries by not building out an infrastructure that is supportive of learning analytics. One example of this is the lack of dashboards or analytical tools at the disposal of students and educators alike. Right now we see the center of the research in this space is primarily focused on the adaption of content. What is lacking is a clear set of actionable steps for both students and teachers on what should be done to get the best possible results. That would look like an increased focus on feedback and reflection for both the student and the instructor as well as allowing for effective intervention from the teacher to the student.

We see this beginning to shift in recent years with renewed discussion on how we can shift towards more data driven discussions in the education sector. However, a considerable hindrance to this remains that there is debate regarding the extent to which data collection should occur and the ramifications of sharing the data.

A critical driver in the growth of the field has been the increase of curriculum available and delivered online to students. One example of this is full courses that are available online. The first surge of involvement in the space from a student perspective can be seen in 2012, when there were over 6.7 million students enrolled in at least one online course [1]. As of 2016, the number of students pursuing higher education who were enrolled in an online course was over a quarter [12].

In addition to classes being offered and made available online, we also have that student assessments and assignments are being moved to a digital platform. This is a critical step towards fostering an environment that not only allows but encourages the personalization of education made feasible through learning analytics.

This is occurring in both elementary school and higher education settings. For younger children, a system such as i-Ready provides a platform for differentiated learning. What i-Read is able to do is that it "integrates powerful assessments and rich insights with effective and engaging instruction in reading and mathematics to address students' individual needs" [24]. In institutions of higher educations companies such as Canvas provide an environment for students to interact with material from lecture recordings to quizzes through a single interface. On other other side of this, all this data can be collected in one place.

Performance Metrics:

When looking at the impact of learning analytics, there are certain metrics that are considered noteworthy within the industry. Administrators often look to student retention and graduation rate numbers to see the impact of a change in the existing organization or process. These are numbers that are typically used as benchmarks in the education space and, as such, have carried over into the education technology space as well. This is particularly true when the focus is on a big picture view of the organization and can be used to compare across geographic regions.

With learning analytics, we additionally have the consideration of learners. Here we see that another important metric is introduced, one that is more applicable at an individual level. This is the extent to which an individual student is understanding the material and engaging with it. While this is the overall goal that is being worked towards, the manner in which it is measured or studied varies a lot depending on the application. Examples of this include the speed and accuracy with which an assignment is completed, the frequency of engaging with the material and the extent to which one contributes to discussions online.

Looking at the same situation from the instructors perspective, we have action research. This is where "through iterative cycles of action, perception and evaluation, teaching can regularly be matched and adjusted to all learners' needs" [6]. The goal of action research is to have steps you can take to have effective practices [1]. What is happening is that there is an active change made in an organization (often a classroom in the education technology space), and the effects of that change are observed. This in theory provides insight into effective versus not effective teaching practices. These insights can then be used to shape what is taught in the classroom and the manner in which it is taught.

Stakeholders:

There are a lot of different individuals who are considered to be stakeholders in the learning analytics space and they all have different objectives that they want out of learning analytics. Their differing goals and perspectives, when not properly addressed, create hindrances to the adoption and support of learning analytics by key individuals. Below are some of these key groups as well as their goals with learning analytics[6].

- Teachers
  They want to understand how they can best communicate their material to the students.
- Students
  They are focused on retaining information and improving their grades.
- Administration
  We see that administrators often make decisions with student retention, graduation rates and identifying students who are at risk at the forefront of their mind.

If learning analytics is used to address these concerns, the next consideration is to what extent existing processes would have to be altered for these to be addressed? Would it simply require an additional layer of work on top of what is being done? Or do we have to alter our fundamental approach in order to be able to reap the full benefits that learning analytics has to offer?

As learning analytics opens the door to new ideas, there are potential clashes that could occur amongst stakeholders in terms of what factors should be considered in making the optimal decision. An example of this would be if the administration wanted to know which teacher is using the most effective teaching method. The teachers could feel as though they are constantly being assessed through learning analytics and feel that they don't have complete freedom and control in their classroom. Another potential clash of conflicts is if student data is used to better inform teacher decision making, but does so in a way that is harmful to the student.

Applications of Learning Analytics

Objectives:

There are a range of situations in which learning analytics can be used and their effects vary across the stakeholders. Firstly, it can be used to monitor and analyze the student's progress. Secondly, learning analytics provides a way to effectively predict and intervene. It gives you the ability to see how the student might be doing in the future and what can be done to keep them on track or, if they are headed down a path that is not positive, what can be done to stop that from happening. Another objective is tutoring and mentoring which is more specific to a certain topic. Learning analytics can be employed for assessment and feedback as well as adaptation. Advancing in this direction, learning analytics also serves as an integral part of personalization and recommendation. This is the objective that is most focused on the learner. In personalization and recommendation there is a back and forth between the knowledge push versus the knowledge pull. Who is leading the way that the learning is happening—the instructor or the student? Another key objective of learning analytics that at times is not as evident in the applications is reflection.

We should draw upon the performance metrics used in the education technology space to evaluate how a given learning analytics platform performs with the objectives listed above. Traditionally, in the education space we have that grades, graduation rates and student retention are the primary factors considered in performing an evaluation. With the introduction of learning analytics, we can expand this to consider other facets of the space such as the rate at which a student is making progress and how impactful a shift in curriculum or a student's study habits are on their ability to understand material as judged based on assessment and feedback. Learning analytics allows us to consider the more individual level performance metrics associated with education.

MOOCs:

Massive Open Online Courses or MOOCs serve as a springboard in the learning analytics space. MOOCs have grown in popularity in recent years. In understanding the role they play in learning analytics it is important to understand what MOOCS are. Massive relates to the number of individuals enrolled in the courses. Open indicates the high level of accessibility of the courses. Online shows that the classes are on the internet. Courses refer to the fact that material is structured in a manner that is meant to be taught to someone else. The advantages of MOOCs are that they can bring education at a scale and depth that previously could not be comprehended at its low cost.

MOOCs provide an opportunity for companies in the learning analytics space to take a rich dataset and begin to scratch the surface of the insights learning analytics has the potential to provide. This in turn allows us to alter the way in which the curriculum is delivered to the learner.

One of the biggest issues faced by MOOCS is their retention rate, and considering what can be done to encourage individuals to complete a course they have started is a one of the problems learning analytics is looking to solve. Just as in schools, there are students in MOOCs driven by both intrinsic and extrinsic motivation. Students are intrinsically motivated when they sign up for a class because it covers a subject matter they are interested in. On the other hand, a situation involving extrinsic motivation in a MOOC would be when an individual has enrolled in a class because they desire the certification they get at the conclusion of the course[16].

In an effort to garner the most applicable and generalizable insights about students that are partaking in MOOCs, the learning analytics clustering approach is often done in this space by the type of student relating to their level of engagement. One way in which that has been done is the portion of the curriculum that the student has worked their way through. Khalil and Ebner considered the following characteristics of students in a study they did in 2016: “completers” who finished the entirely of the course; “auditors” who watched the videos but did not do some of the assignments; “disengaging students” who left the class after the first third of it; “sampling” who did not continue with the material after the first two weeks [16]. Another way in which students can be divided is the extent to which they completed assignments.

In a study that looked to how students were motivated in MOOCs that focused primarily at differentiating between intrinsic and extrinsic motivation, the results had strong indications for future research on how to reduce dropout rates in MOOCs. The findings revealed that the four main groupings of students mirrored those of Cryer’s scheme of Elton which looked at a commitment of an individual in a classroom [16].

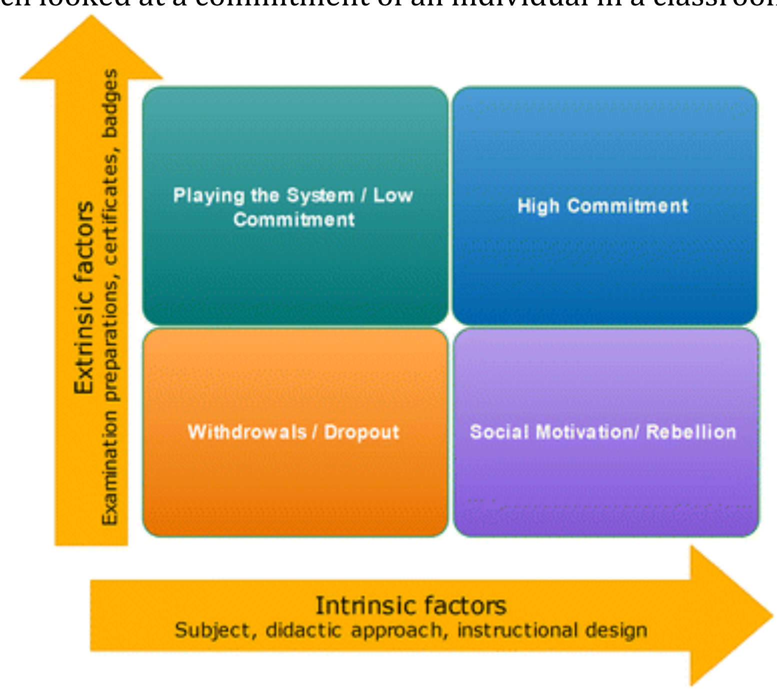


Figure 4: Cryer’s scheme of Elton
Source: Clustering patterns of engagement in Massive Open Online Courses [16]

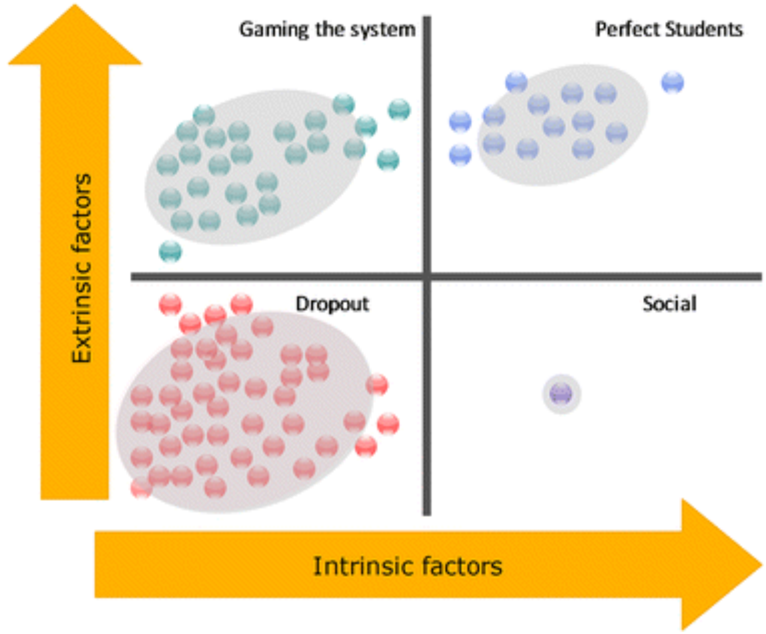


Figure 5: Clustering analysis of students involved in a MOOC
Source: Clustering patterns of engagement in Massive Open Online Courses [16]

Observing the clustering that came about when considering students that were taking a MOOC, the results of the cluster mirror Cryer's scheme of Elton. This draws a strong link between the profiles of students in the classroom and those using MOOCs to be in some core ways analogous. This gives us insights on how to shift students into the types of students that gain more from an online course. The results of the study pointed towards increasing intrinsic factors as opposed to extrinsic factors as a way to involve students [16].

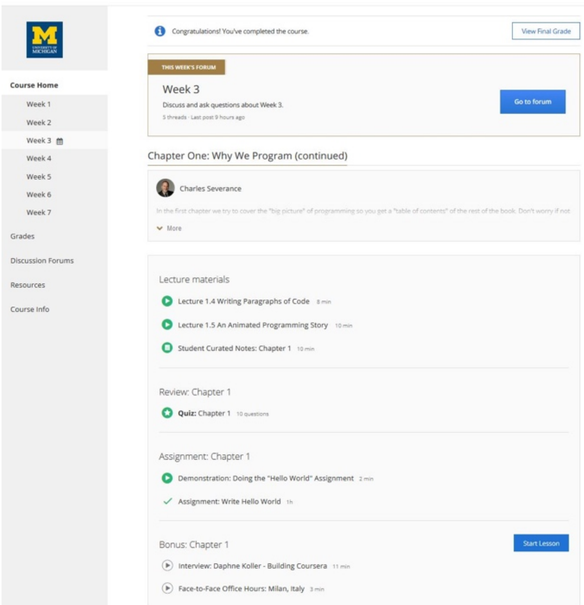
Case Studies:

As we have discussed the space of learning analytics, we have developed an understanding of the potential that this technology holds. In recent years, there has been a rapid increase in the number of education technology firms and the field has become increasingly saturated. Since 2011 we have seen a continuous growth in money put into education technology companies, and for the United States that number reached a peak in 2018 hitting $1.45 billion [25] . There are a lot of different types of companies in this space, many of them addressing different niche markets within the space.

As is often the case, there is a gap between the potential of the technology that is considered in a purely theoretical and research context as opposed to the actual technology that a company builds a business model on. Here we will explore three companies — Coursera, Duolingo and Signal. Each of these firms is at the forefront of the problem they concentrate on solving, and as such give us an understanding of the breadth of the field of learning analytics as it currently stands.

Coursera

Coursera is an example of platform for massive open online courses or MOOCs. The platforms offers online courses the cover a range of topics from "Managing Innovation and Design Thinking" to "Child Nutrition and Cooking." Coursera partners closely with universities to have highly qualified instructors curate the material and teach the courses. The company was founded in 2012 and has grown to more than 30 million users in 2018 [26].



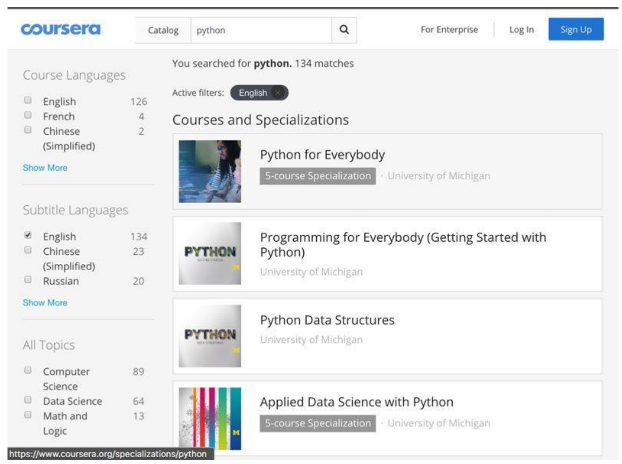


Figure 6: Example Curriculum for Course
Source: Coursera

Figure 7: Sample Courses Offered
Source: Coursera

The videos for the classes being taught are available for free. There are a few streams through which Coursera generates revenue. To begin with, to get a certificate showing that you have completed the material associated with the class, the consumer pays a fee. There is an additional fee for grades and assessments, which goes to show the validity with which you have engaged with the material of a given course. Most recently in 2016, they have added a monthly subscription model that allows one to get a specialization [26].

One nuance to consider is that a MOOC such as Coursera offers courses in a range of different subject matters. As such, one must keep in mind how engagement and analytics of engagement might vary from say English to mathematics.

Coursera plays a unique role in the larger field of data analytics, as it is a source of one of the largest public data sets relating to students and learners studying a curriculum from material online. We can look at data from Coursera MOOCs to understand how people are engaging with online course work. The data that is being considered when studying learner behavior on Coursera includes information such as where they click, what they view, how long they are on a page etc. It has been found that there is an exponential decay in the individuals actively in a course throughout the first 6/10 of the duration of the course.

As this data set is further studied, we want to be able to answer the following questions:

- How do students interact with the online material?
- Can we gauge from the data we can collect when a student is enrolled in a MOOC whether or not they know the material?

This has serious implications for the validity of certifications that MOOCS have started to give. In order to better understand this, the process through which students engage with the material they are presented in a MOOC is studied. What has been discovered thus far in regards to the processes that allow students to be successful in a MOOC is that successful students watch videos in the order recommended by the instructor.

We want to be able to consider less of the demographic information of the student and more what their habits are when they are learning through a MOOC and use this to serve as the baseline for how they are learning. This allows for less biased as well as more generalizable conclusions. In order to be able to do this, process mining is used.

Process mining involves three steps. The first part is to create a model, which is called discovery. The second part is conformance, which is when you compare the model that has been discovered to the data and determine where in the deviations lie. The third part of process mining involved enhancement, which is where the model is altered based on what was potentially uncovered during conformance [3].

A study done building a process model on Coursera using its data found that the following two factors were integral in understanding learner behavior on these MOOCS: [17]

- Watch status
- Viewing habit

Duolingo

Duolingo is a language learning platform that teaches English speakers 11 different languages. The company has a website as well as an app that has over 300 million users []. Their business model is such that they don't charge consumers who utilize the application. Instead they derive their revenue primarily through advertisements. If a user wishes to utilize the application without advertisements, they can pay a subscription fee for an ad free service.

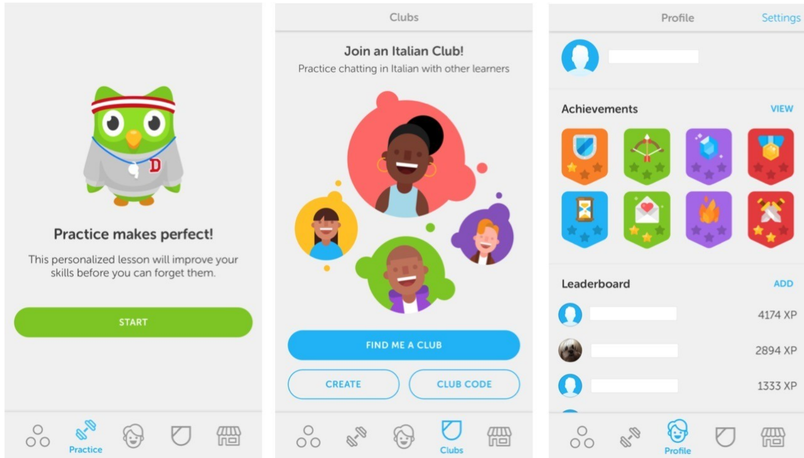


Figure 8: Introduction to a language community in Duolingo
Source: Duolingo

Let's consider the impact of Duolingo through a study that explored the effectiveness of Duolingo in learning Spanish as measured by a score on a college Spanish placement exam. In the study, learners took the placement test twice, once before and once after they worked with Duolingo. Overall, the conclusion was drawn that Duolingo was successfully able to increase the learners' understanding of Spanish[29].

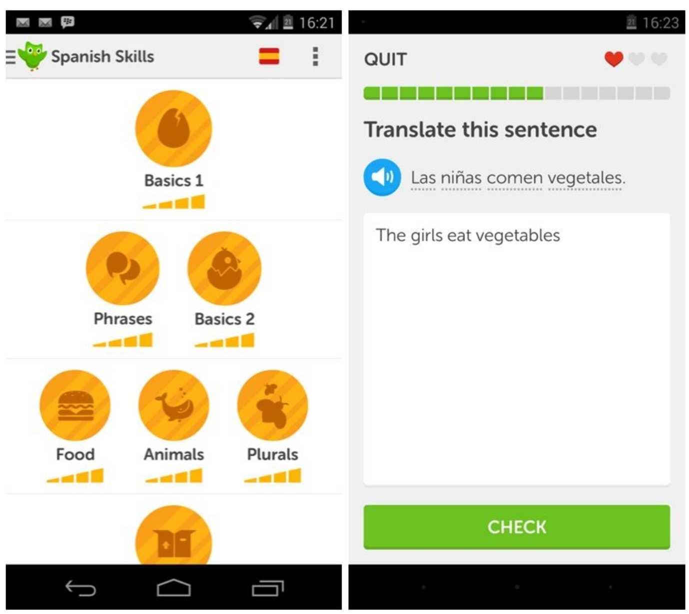


Figure 9: Interface to learn Spanish in Duolingo
Source: Duolingo

The factors that played into the success of the subjects of the study revealed an inverse relationship between how much the learner initially knew about the language and how much gained from Duolingo. When studying the impact of Duolingo in this setting, one cannot just look at the increase in the score in the two times the exam was taken. The amount of time the learner spent on Duolingo must also be taken into account. So, in these situations the metric of points gained per hour of study is a common metric used to measure the success of the online learning. The overall sentiment of the users on the technology was that they were satisfied with it and didn't have a difficult time learning the subject matter through a non traditional avenue. This held true across age groups, levels of education and racial backgrounds.

Duolingo relies heavily on adaptive learning. The big picture idea of adaptive learning that Duolingo assumes is that the material the learner is presented with is modified based on their past performance. This is done using a decision tree model.

The idealist goal with this is a personalized education system, which is not possible in the existing classroom setting. So you have a space where companies like Duolingo can add a lot of value.

But an important factor to consider when teaching material through a decision tree model is that learning a language through adaptive learning has significant differences from learning a subject such as mathematics. Adaptive learning is based off of two key assumptions. The first assumption is that the material is being learnt in a linear fashion, and, as such, that it is cumulative. The second assumption is that the material can be broken down into building blocks that allow one to be taught the material in pieces [3]. Although this makes sense for a subject such as mathematics, this doesn't necessarily translate over to languages and as such it becomes significantly more complex to apply an adaptive learning model in teaching someone a new language. A large part of this is how the extent to which one "understands" or has "mastery" of a subject is evaluated.

Signal

Signal was first piloted in Purdue University in 2006 and started being used in a multitude of other settings in 2009. It is a platform that monitors a student's progress and alerts them in real time if there is an area in which they should be doing additional work. If they receive a "signal" indicating that they are not on tack, then their professor can provide them information on what they can do to get back on track [13].

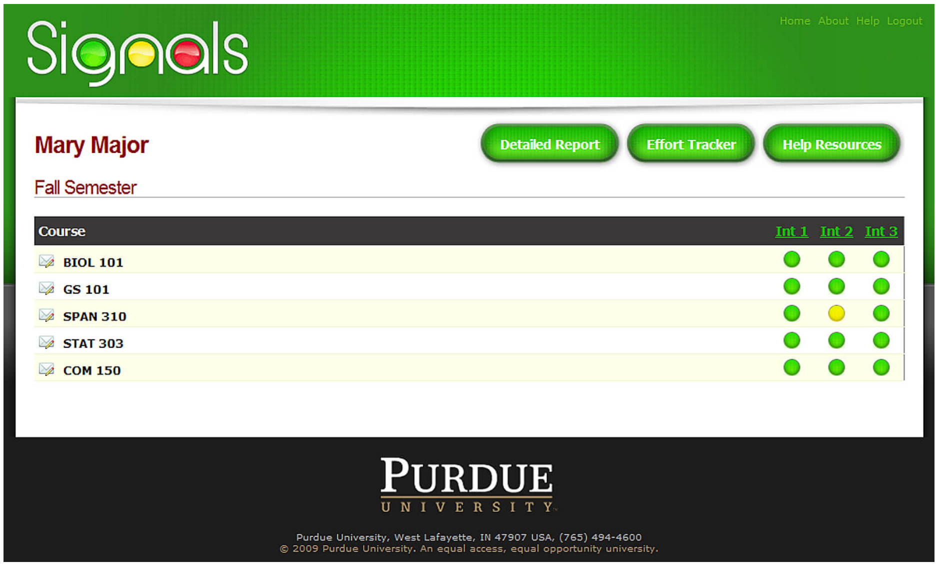


Figure 10: Sample interface of Signals application
Source: Signals, Purdue University

In looking at how the system works, we see that it is not just taking into account the actual grade the student has. Information is also used on how the student has engaged with the material they have been looking at. It essentially uses this information as well as their actual scores and then compares it to previous students to gauge how this student is performing.

The key pieces of information that go into the algorithm that Signals system uses are listed below:

- "Performance: based on points earned on the course so far
- Effort: interaction with the VLE compared with other students
- Prior academic history: including high school GPA and standardized test scores
- Student characteristics: e.g. age or credits attempted" [29]

The impact that Signal has on students is that it helps them realize they need to improve and what they need to do to improve before it is too late for them to catch up. It helps to overcome the issue of not being able to help students that fall in the middle of the class (as opposed to focusing on the top or bottom tier students). The results of utilizing Signal showed that there were 10% more A's and B's given in classes that had Signal. When students were no longer in the Signal program, the students that had used the platform were still more likely to get help outside of class and make sure they were keeping up.

Data Privacy, Regulation, and Ethics

Until now we have examined what the learning analytics space looks like from the perspective of what the potential of the technology is and how it is currently being used in industry. An additional dimension to assess is the role that data privacy and regulation plays. The main sections of this are policy, corporate social responsibility, and public opinion.

This is one of the largest implications of the increase of technology in the classrooms. With this introduction of technology into the classroom space, there is significantly more information collected about students. Now, since this information is electronic, it provides the incredible benefit of researchers and technologists being able to pool the data for educational data mining. On the flip side, there is the potential for the data being used with mal intent as well as it spreading more easily without the individual who created the data having control over it. This sparks a deep controversy about the extent to which this data should be collected to begin with.

Regulatory Landscape:

The education technology space can be molded from the insights of learning analytics to create an environment where each individual has the opportunity to reach their maximum potential. This is what the educators have striven towards for hundreds of years. A significant hurdle for this to become a reality is the regulation that results in walls being built around student data that can make it particularly difficult to parse for insights. There is legitimate cause for their to be concerns around who has access to what data about a student, particularly considering the role that age plays here. Understandably, there is a lot of apprehension that children cannot sufficiently perceive risks that might be associated with decisions that they are making and, as such, could be coerced into doing something they aren't necessarily comfortable with.

Two main existing laws that are connected to the rights that minors have in regards to privacy in the education and technology space are the Federal Education Rights and Privacy Act (FERPA) and the Children's Online Privacy Protection Act (COPPA).

- FERPA:

  The Federal Education Rights and Privacy Act is limited to organizations that receive federal funding in some capacity. If an organization violates FERPA, there is a financial penalty that they then must pay. The act relates to educational records which encompass the following buckets of information: "According to FERPA, education records contain information on student background, academic performance, grades, standardized test results, psychological evaluations, disability reports, and anecdotal remarks from teachers or school authorities regarding academic performance or student behavior" (FERPA, 1974, 20 U.S.C § 1232g (a)(1)(D)(3) [18].

If this information is to be released by the school, then they have to get written permission. If the student is under the age of 18, then the release must come from a parent or guardian over the age of 18. If the student is over the age of 18, then they can provide the written consent themselves. When this information is passed along to another organization, in compliance with FERPA, it is under the understanding that the information shared will not go beyond that.

There is a gray area and debate around a few areas in FERPA. Do emails that are hosted by the school's cloud services get protection under FERPA? What if data is being collected to be utilized for some sort of additional service?

- COPPA:
  - The Children's Online Privacy Protection Act applies to youth under the age of 13. Under COPPA, online providers who are collecting data from children under the age of 13 must garner parental consent that they can collect the data about the child. It is important to note that with this act, parents also have the right to, at any point, retract the right for the provider to store this information. If they decide to do so, than any previously collected data must also be deleted [18]..

  - COPPA was most recently updated to include the following items as also being considered personal information for which organizations must receive consent: cookies, geolocation, photographs, videos and audio recordings. This is a list that has been expanded and updated over the years as the number of ways in which data about individuals of this age can be collected has grown.

Privacy Concerns:

As with any new technology that emerges, we must consider the larger implications that it poses for society. With learning analytics, there are a plethora of concerns around the potential misuse of data that is being collected as well as tension around who has ownership of the data and who is privy to the information that is being collected around this.

There is a general consensus that there are benefits to the personalization and adaptability that the platforms that have come about as a result of learning analytics have to offer. But to be able to collect the data to perform the analysis, companies find themselves pulled in multiple different directions addressing the questions mentioned above.

The overarching concerns regarding privacy of student data in learning analytics can be broken down into the following five buckets[18]:

- *Issue One:* The network of people and organizations that have access to the information. If we allow a company or an educational organization to collect data about a student, does this make it more likely that the data will be passed along to another entity?

  Here we face the issue of if a person gives permission to Organization A to have their data, what happens if Organization A then passes it along to Organization B? There are a multitude of complicated situations in the learning analytics space that can arise as a result of this. For example, say data is being collected on a learning management system on how many videos a student is watching and what portion of the video they have viewed. There are studies that indicate that this is correlated to student performance in the course and is therefore relevant information to retain. But just because this information is beneficial to the LMS to more accurately determine a student's success in the course does not mean that the instructors should be privy to this information. It could result in the teacher being biased for or against the student.

  Another case to consider is one where the argument could be made that information should not be shared with the student. Referring back to the earlier discussion on Signals, in giving a student a red light, while this is helpful information in determining where a student stands, it can also result in the student feeling hopeless and unable to overcome the task ahead of them. By providing them with this information, we run the risk that they if they didn't know this, they would have kept working hard and improved their score. But now they might just write themselves off as not good enough and stop working towards success in that class or subject.

  There is also the consideration of how information is shared with third parties. Student records are strictly protected under FERPA. However, consider the fact that prospective employers already ask students for information such as transcripts and letters of recommendations, which the students then allow the release of. Although analytical information gathered about a student would be protected under FERPA, if this information existed, there is a high chance that employers would want to be privy to it. The intensity of the competition around jobs would most likely result in a scenario where students don't have much of an option other than to hand over this information to employers. There are those who argue that it is beneficial to the efficiency of the economy if information about individuals is made open and this in turn results in greater social welfare [9]. This example addresses the fact that the data collected with the purpose of helping an individual learn isn't necessarily only going to be used for that purpose.

- *Issue Two:* How intrusive is the process of gathering information? If there a point at which the information being collected is beneficial but too intrusive to be stored?

  When determining what information to collect, the primary consideration is whether or not it is relevant. In learning analytics, we are looking at a field that has relatively recently existed in its current form and as such almost anything could be considered to be potentially relevant. Individuals who support this approach are prioritizing the insights gained from learning analytics above all else [10]. This could then be used as justification to collect information about all aspects of an individual's life ranging from their political views to their religious beliefs.

  The big question that is posed here is if increasing a learning outcome is the only factor that organizations in this space should consider when gauging what information they should and should not collect. This is an issue that is particularly relevant given that data can also be sources from things like social media and geo location which would provide schools and universities with the above mentioned information.

- *Issue Three:* Weighing the upsides and downsides of the outcome of insights from analytics to determine if it is justified. Taking into consideration the risks, consequences and the manner in which the consequences are distributed, does that at any point outweigh the benefits gained from learning analytics?

  There is substantial work done that points us in the direction of supporting the idea that learning analytics provides additional insights about a learning environment which allow institutions to adjust their structures to better tailor the experience. What we are addressing here is whether this goal justify any unintended consequences? This becomes particularly relevant in situations where we consider that the privacy concerns of a student diverge from what is in the best interests of the institution. An additional factor to consider is if the benefits of learning analytics are not evenly distributed amongst students.

  An integral component to this discussion is a student's right to privacy and the repercussions of having a student's priorities diverge from that of the institution. We have discussed heavily the advantage that learning analytics provides in terms of guiding instruction in the classroom and determining how a student would be most successful. Say that we were then able to create a list of classes a student should take in order to be most successful. Now, we are guiding students into taking classes that they otherwise might not have, which isn't necessarily in the benefit of the student but instead the institution at large which is working to optimize on a few metrics.

- *Issue Four:* Impact on a student's autonomy. Do the benefits from learning analytics outweigh a student's say in what information is being collected about them?

  Here we are faced with the dilemma of how much control an individual should be able to have over data collection for learning analytics. Given that it is still a relatively new field, there are is not a clear list of ways in which the data you are allowing to be collected about you is going to be used.

  In addition to the data being collected about them, there is another element of whether students should have control over what the data is being used for.

- *Issue Five:* Does the work align with the goals of the education system?

  Relative to the other issues, this one addresses a much broader viewpoint. At the core of learning analytics is the idea that we are working towards a betterment of achieving learning objectives. So in discussing this issue, we then must consider what the learning objectives are. To define it as good grades for students would be narrowing the scope of it to less than what it is. Additional learning objectives include communication skills, the way in which one approaches a problem, being an active and knowledgeable member of one's community etc. Given that the goals are such broad concepts, it becomes difficult to assess to what extent they have been achieved.

  A common benchmark used in institutions of higher education is the statistics around job placements. Given that the information around job placement is most easy to measure and feed into a model, there could be undue importance given to this in learning analytics. An example of this going wrong would be if classrooms full of individuals were pushed to take a class that wasn't necessary because it made them more employable.

At the heart of a lot of these issues is that, until learning analytics emerged as a more prominent discipline, there was a clear line between data used for assessing a student and a student's personal information such as where they live and how they work. As learning analytics grows, this line continues to blur.

There are ways in which to accommodate to mitigate and decrease the risks and issues brought up above. By being mindful of the ripple effect of learning analytics, systems and policies can be put in place to avoid scenarios such as the ones mentioned above. Factors to consider in addressing the issues are mentioned below:

- *Addressing Issue One:* Be conscious of all possible implications of sharing information with a given stakeholder. While it might be beneficial to share information with one party, it could be detrimental to share the same information with another party. As such, the information garnered from learning analytics should always be set up in such a manner that it allows for different degrees of access. Remember that individuals and organizations outside of the education space might eventually be privy to the information being collected.

- *Addressing Issue Two:* A step in drawing a line about what data can and cannot be collected would be to establish criteria for determining this beyond potential relevance. We have the basis that data is being collected to further the learning experience of an individual. In order to make it better, one has to be able to change something so that the outcome can be different. One potential criterion is that the institution should only collect data on things that they can actually change and help the student with. Thus if it is deemed that an educational organization should not intervene on the basis of religion or political views, then it should be impermissible for such data to be collected to begin with.

- *Addressing Issue Three:* Relative to the other issues, this one is more nuanced to address. This is because, in order to get guidance on the best course of action, there has to be a decision made on what is considered to be just in terms of the benefits and consequences.

- *Addressing Issue Four:* There is a strong argument to be made here that creating an environment of transparency would be foundational in trust between individuals and institutions collecting data about them. If individuals felt a sense of control over the data that they were sharing, this would potentially make them more open with allowing the data to be collected.

- *Addressing Issue Five:* It is impossible to look at this issue without considering the other four problems defined above. To know whether learning analytics is an that is worthy of the comprises is a large question to ask. How much one values self autonomy and control over larger insights creates a tug of war within this industry.

There are a range of concerns that are at the forefront of discussion around data privacy issues in regards to data analytics. The argument has been made that the data collected is significantly more beneficial to the institution that is collecting it then it is to the student [1]. Furthermore, although big data can be seen as adding a transparency to the world, there is also an issue of the decisions that are being made as a result of big data being seen as coming out of a black box. In the context of learning analytics, this means that individuals have little knowledge of what information is being used to come to what conclusions about them[11].

As mentioned above when discussing the types of data that are pooled together for learning analytics, a student's demographic information, socioeconomic status, gender identity, sexual identity etc. could be used. As a society that has been making prejudiced decisions for generations, we run the risk of the conclusions drawn from this data mirroring old and existing prejudices. This would in turn result in systematic biases. An additional consideration is that as people become more aware of the data that is being collected about them, they alter the manner in which they are acting and this influences the conclusions then drawn from the data. This is referred to as the chilling effect [7].

Implications of Privacy Concerns:

In considering the privacy issues faced by the learning analytics space, we can look to other industries in determining best practices and standards to use as a benchmark. One industry in particular where privacy concerns strongly dictate how data is shared is the healthcare space. This is an example of there being heavy regulation that sets the roadmap for what is acceptable and what is not acceptable. On the other hand, when looking at the commerce space, there is considerably less regulation on what a company can do with the data they collect and also who they can share it with. However, a combination of data leakage issues as well as growing consumer awareness in regards to the mass amount of data being collected about them has resulted in a push for more regulation in this space as well. It is important to note that none of the regulation, not even that in the healthcare space, points towards "absolute confidentiality"[8].

In order to have stricter regulations, there is protection of data that could be individually identifiable. When the data cannot be linked back to a specific person, the rules of sharing it between organizations and the purposes for which it can be utilized are looser.

When discussing privacy issues, we draw a distinction with age and there is a significant difference in the policy for minors versus individuals over the age of 18. The question then arises if this is a fair place to draw a distinction? Given the reduced amount of regulation in collecting data on adults, it is significantly easier for firms to do so, and as such they have more data about adults available to them. Since it is safer to solicit permission from adults to get data, is there a potential to regress backwards from data about adults and then use this to build models for students?

If the sector is less regulated in terms of data usage, there are situations in which institutions take it upon themselves to use the data responsibly. This is seen as corporate social responsibility dictating the way in which data is used. Although the educational sector is not one that is lightly regulated, particularly in relation to learning analytics, the idea of corporate social responsibility still plays a key role here. Since the inner workings of learning analytics is seen as a black hole to most consumers, in order for parents to sign off on their children's data being collected, they have to believe that the data will not be used and the results will not be detrimental to their child's learning experience. Corporate social responsibility in regards to learning analytics can come in many different forms: making sure the storage of the data is held to a high standard of safety; anonymizing the data whenever it can be done; not sharing data with third parties; open communication with the consumers on what data they are collecting and to what end it is being used.

Data as a Commodity:

There is another facet to this discussion which takes into account how data is viewed in other sectors. In a multitude of consumer facing industries, we see that data has a high monetary value and is sold by companies as a significant portion of their revenue. This idea views data as a commodity. Looking back on the type of companies that we saw in the learning analytics space, this is not a source of revenue for them given the strict limitations of their data usage.

Given that they are constrained in their ability to monetize directly from their data collection, this could in some capacity deter them from enlarging the data that they are collecting. Considering the revenue stream for a lot of companies in the analytics space is closely tied to their data collection, this is a considerable constraint for some companies in the the learning analytics space.

An example of this was revealed in an interview with Rob Richardson, the CEO of Pocket Points [30]. Pocket Points has a phone application that gives students points for having their phones off during class. These points can then be redeemed for rewards such as extra credit or a coupon for a local store. The goal here is too incentivize students to stay off their phones during class. However, the firm struggles to maintain their profit margins and one good solution would be to have advertisements within the app. However, given the regulations of the space, they are unable to get enough data on a student to provide targeted advertising. As such, companies don't think it is worth their marketing money to advertise through Pocket Points app. This is not to say that they should be allowed to advertise to students, but rather to draw attention to the complications that arise in the field when data cannot be viewed as a commodity.

Conclusion

Although regulation should serve as a way to safeguard students' rights and ensure that they are not put in harm's way, it is incredibly beneficial to both them and schools at large for their data to be more openly shared. Without having access to large pools of student data, companies in the space are not incentivized to invest in growing the scope and applications of learning analytics. All three key stakeholders stand to benefit from learning analytics becoming a larger part of educational organizations.

From a student's perspective, the benefit lies in an individualized curriculum. The two primary advantages of this are that we can gravitate away from the traditional classroom model of putting more effort towards the students at the top and bottom of the grading distribution and provide a platform that supports students at all levels. The second stakeholder, which is the teacher, can gain a lot from knowing what the most effective components of their teaching practices are. They can then further hone these skills and develop a more targeted effort towards making the most of the time they have with their students. Considering the third stakeholder, the administration and larger institutional goals, their first benefit comes from students and teachers performing better. Having students get the best education they can is reflected in an increase in standardized scores that are often used at performance metrics to make sure a given school is up to par. Furthermore, it provides them the capacity to streamline processes within the institution, resulting in greater operational efficiencies.

All of the above mentioned potential growth that can come out of learning analytics is contingent upon the organizations and individuals driving this technological advancement having access to student data. Herein we see complications begin surrounding access to data, particularly about minors. This can create serious barriers for the technology to be able to reach its full potential. This becomes an even more complicated matter when we consider that companies in this field might be disincentivized to collect data since they do not seen how they can build a profitable business model with this constraint.

Learning analytics has the potential to make a positive impact on the future of humanity in a very fundamental way. The rapid advances and huge investments in AI, ML and other technologies are poised to deliver powerful algorithms and techniques that will make analytics more insightful and actionable. The proliferation of digital devices will continue and the amount of data available to analyze will also continue to grow exponentially. Technology will exist where it will be possible to identify, intervene and influence a student's learning experience so that it is hyper-personalized and brings out the best individual outcome for that student. This can be made available to students early in their education and regardless of their socio-economic status. The positive implications of such a change are manifold. The regulatory environment will need to adapt and ensure that it is enabling such a change instead of being an element of friction. That will determine the impact of learning analytics more than any other aspect.

Learning analytics is faced with a catch 22—we cannot be sure what the risks and rewards of learning analytics applications look like at the scale we envision them to be until we have enough data to put it into action. But we cannot allow the data to be collected until we are able to weigh risks and rewards and shape regulation accordingly. So as companies pioneer into this landscape, it is critical that they keep at the core of their learning analytics efforts their sense of corporate social responsibility and hold their data practices to a high standard. This is critical in building trust with the other stakeholders, and it is only through the support of all three stakeholders that the learning analytics space will be an environment that fosters growth for these companies.

We can look forward into what a classroom might look like 5 years from now: desks that are interactive, computers that provide personalized assessments, a virtual reality corner where students practice debating. Learning analytics has a key role to play in making this a reality. As technology steam rolls through the current classroom, creating an entirely new environment that we didn't even know we could aspire to, we cannot allow the ethical implications of this to lag behind. We cannot forget that the individuals most impacted by this are often ones too young to understand the implications of the change and have a say in it. Their interests and safety must be at the forefront of discussion as we build our future classroom.

Endnotes

[1] Rubel, Alan P. and Jones, Kyle, Data Analytics in Higher Education: Key Concerns and Open Questions (May 2016). University of St. Thomas Journal of Law and Public Policy, 2017. Available at SSRN: https://ssrn.com/abstract=2905985

[2] Honey, Kristen. "My Data: Empowering All Americans with Personal Data Access." *National Archives and Records Administration*, National Archives and Records Administration, Mar. 2016, obamawhitehouse.archives.gov/blog/2016/03/15/my-data-empowering-all-americans-personal-data-access.

[3] Baker, Ryan Shaun Joazeiro de and Paul Salvador Inventado. "Chapter X : Educational Data Mining and Learning Analytics." (2014).

[4] Omedes, Jose. "Learning Analytics 2018 - An Updated Perspective." *IADLearning*, 15 Feb. 2018, www.iadlearning.com/learning-analytics-2018/.

[5] Yudelson, Michael V., et al. "Individualized Bayesian Knowledge Tracing Models." *Lecture Notes in Computer Science Artificial Intelligence in Education*, 2013, pp. 171–180., doi:10.1007/978-3-642-39112-5_18.

[6] Mohamed Amine Chatti, Anna Lea Dyckhoff, Ulrik Schroeder, and Hendrik Thüs. 2012. A reference model for learning analytics. Int. J. Technol. Enhanc. Learn. 4, 5/6 (January 2012), 318-331. DOI=http://dx.doi.org/10.1504/IJTEL.2012.051815

[7] Solove, Daniel J. "A Taxonomy of Privacy." *University of Pennsylvania Law Review*, vol. 154, no. 3, 2006, pp. 477–564. *JSTOR*, www.jstor.org/stable/40041279.

[8] Jianfeng Xia, Min Yu, Ying Yang, Hao Jin, "Personalized Privacy-Preserving with high performance: (α ε)-anonymity", *Computers and Communications (ISCC) 2018 IEEE Symposium on*, pp. 00807-00812, 2018.

[9] Posner, Richard A. "An Economic Theory of Privacy." *Philosophical Dimensions of Privacy: An Anthology,* edited by Ferdinand David Schoeman, Cambridge University Press, Cambridge, 1984, pp. 333-345.

[10] Diaz, V., and M. Brown. 2012. Learning analytics: A report from the ELI focus session. ELI report ELI3027.

[11] Slade, Sharon, and Paul Prinsloo. "Learning Analytics." *American Behavioral Scientist,* vol. 57, no. 10, 2013, pp. 1510–1529., doi:10.1177/0002764213479366.

[12] “Report: One in Four Students Enrolled in Online Courses.” *OLC*, 25 Feb. 2016, onlinelearningconsortium.org/news_item/report-one-four-students-enrolled-online-courses/.

[13] “Signals Tells Students How They're Doing Even before the Test.” *Signals Tells Students How They're Doing Even before the Test*, www.purdue.edu/uns/x/2009b/090827ArnoldSignals.html.

[14] Siemens, George. “Learning Analytics.” *American Behavioral Scientist*, vol. 57, no. 10, 2013, pp. 1380–1400., doi:10.1177/0002764213498851.

[15] Bogarín, Alejandro, et al. “Clustering for Improving Educational Process Mining.” *Proceedins of the Fourth International Conference on Learning Analytics And Knowledge - LAK '14*, 2014, doi:10.1145/2567574.2567604.

[16] Khalil, Mohammad, and Martin Ebner. “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*, vol. 29, no. 1, 2016, pp. 114–132., doi:10.1007/s12528-016-9126-9.

[17] Mukala, Patrick, et al. “Learning Analytics on Coursera Event Data: A Process Mining Approach.” *Department of Mathematics and Computer Science Eindhoven University of Technology, Eindhoven, The Netherland*, 2015.

[18] Sabourin, Jennifer, et al. “Student Privacy and Educational Data Mining: Perspectives from Industry .” *International Conference on Educational Data Mining*, 2017, doi:10.31094/2017/1.

[19] Upasana. “Decision Tree | Decision Tree Introduction With Examples.” *Edureka*, Edureka, 22 Feb. 2019, www.edureka.co/blog/decision-trees/.

[20] Ferguson, Rebecca, and Simon Buckingham Shum. “Social Learning Analytics.” *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge - LAK '12*, 2012, doi:10.1145/2330601.2330616.

[21] “Power Your School Operations to Success.” *PowerSchool*, www.powerschool.com/ppc/student-information-system/.

[22] Hartmann, T., Moawad, A., Fouquet, F., Nain, G., Klein, J., Traon, Y.L., & Jézéquel, J. (2017). Model-Driven Analytics: Connecting Data, Domain Knowledge, and Learning. *CoRR, abs/1704.01320*.

[23] Malliarakis, Chris, and Maya Satratzemi. “Integrating Learning Analytics in an Educational MMORPG for Computer Programming - IEEE Conference Publication.”

Institute of Electrical and Electronics Engineers , ieeexplore.ieee.org/abstract/document/6901446.

[24] “Personalized Learning for All Students.” *Personalized Learning for All Students | Curriculum Associates*, www.curriculumassociates.com/products/i-ready.

[25] Wan, Tony. “US Edtech Investments Peak Again With $1.45 Billion Raised in 2018 - EdSurge News.” *EdSurge*, EdSurge, 15 Jan. 2019, www.edsurge.com/news/2019-01-15-us-edtech-investments-peak-again-with-1-45-billion-raised-in-2018.

[26] “By The Numbers: MOOCs in 2018 - Class Central.” *Class Central's MOOC Report*, 21 Mar. 2019, www.classcentral.com/report/mooc-stats-2018/.

[27] Lardinois, Frederic. “Duolingo Hires Its First Chief Marketing Officer as Active User Numbers Stagnate but Revenue Grows – TechCrunch.” *TechCrunch*, TechCrunch, 1 Aug. 2018, techcrunch.com/2018/08/01/duolingo-hires-its-first-chief-marketing-officer-as-active-user-numbers-stagnate/.

[28] Philip Kerr, Adaptive learning, *ELT Journal*, Volume 70, Issue 1, January 2016, Pages 88–93, https://doi.org/10.1093/elt/ccv055.

[29] “CASE STUDY A: Traffic Lights and Interventions: Signals at Purdue University.” *Learning Analytics in Higher Education*.

[30] Richardson, Rob. “Interview with Pocket Points CEO.” 16 Apr. 2019.