Analytics projects for key stakeholders in large scale online learning systems

By Rahul Jakhmola

Thesis submitted in partial fulfillment for the degree of

Master of Science in Data Science

August 2021

APPROVED:

Advisor : Dr. Neil Heffernan
Reader : Dr. Mohamed Eltabakh

Department of Data Science
Worcester Polytechnic Institute

Table of Contents

Introduction

As a computer-based learning platform, ASSISTments helped both educators and students across the country by providing a number of tools to aid in providing immediate feedback, to report meaningful data, and deliver instructional support.

This has particularly been the case over the previous academic year where many schools were forced to shift to remote learning in response to the COVID-19 pandemic; during that time,

- over 20,000 teachers led
- more than 500,000 students to solve
- 30,000,000+ problems within the system.

At this scale, it is imperative to not only support educators, but also provide the infrastructural support to developers, administrators, and other stakeholders who are involved in maintaining and improving ASSISTments. Within this, it is important to understand data needs for these different stakeholders, particularly as a data scientist in the team running the e-learning platform.

There is copious amounts of student and teacher data related to homework assignments which has proved useful in studying aspects of learning, but this information also has the potential to positively affect day-to-day decision making for the stakeholders.In this thesis, three different levels of stakeholders are identified within existing data analytics projects, where the level of data granularity required for analysis increases with each level.

These are:

- The ASSISTments administrative team,
- teachers who interact with the system, and
- students who interact with the web-based tutor to complete assignments.

This project seeks to focus on each of these three types of users to explore how the existing data within the platform may improve their differing experiences in support of improving the system.

Project 1 : Descriptive Analytics

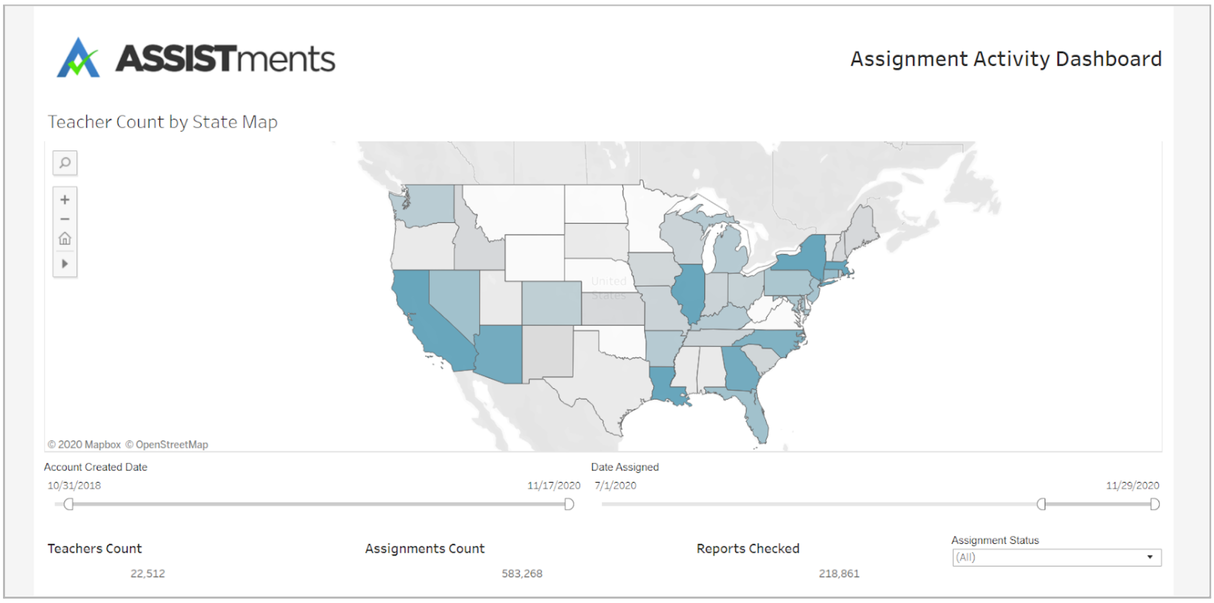
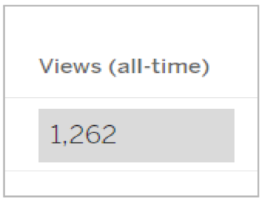


Fig 1 : A Screenshot of the interactive map visualization from the "Assignment Activity Dashboard".



This is the most popular dashboard that has been used more than 1000 times by ASSISTments team members since its first iteration.

The darker blue color is an indicator of a higher number of teachers from states (as of end of November).

0. INDEX

1. INTRODUCTION

The Product, Marketing and Comms teams need regular access to teacher sign up and usage data. This project aimed to develop an **Automated Reporting System that is**

- **reliable,**
- **automated,**
- **user friendly and**
- **robust**.

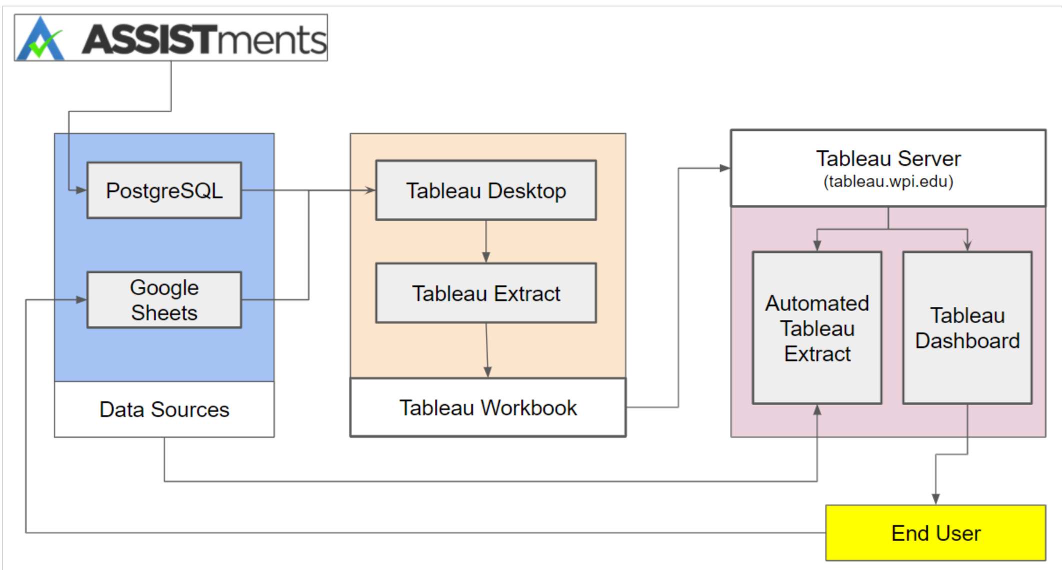
This was achieved through using Tableau dashboards. The teams have been using these dashboards regularly, and a total of 33 Dashboards made till now have been viewed more than 2000 times in this year, with the team members using them regularly for data needs. The questions include, and are not limited to:

- **How many signed up?**
- **How many continue to make assignments?**
- **What content do they use?**
- **How are we doing as compared to last week?**
- **Can we see EngageNY assignment reports for grade 5?**

While this task was being done through SQL queries and pivot tables, it also meant taking a lot of time from getting the data to running the analysis. Also, these queries would need to be rewritten for new features, by different people, and then run by someone else. Also, as the queries were written by different people every single time, the data reliability was low.

2. PROJECT STRUCTURE

The project requires **three components** which would work together to deliver reports that are updated daily - the **Data Sources, Tableau Workbooks and Tableau Server** for hosting the Dashboard. **Fig 2** below shows the path of information among the different parts of the project.



A. DATA SOURCES

Tableau allows for using 100+ types of data sources. However, we use two options that allow reading the data from the ASSISTments dev server and manual entry of data as well when needed - PostgreSQL and Google Sheets.

PostgreSQL (dev.tng)
This is the ASSISTments **dev server**. However, the database has different schemas for ASSISTments 1.0, ASSISTments 2.0 and 2.0 logs. Then, there are different tables in all three schemas that need to be joined to get the required data. The following ER diagrams list the commonly used tables for answering reporting questions for 1.0 and 2.0.

Google Sheets
For security purposes, the ASSISTments google drive was used to host tables in google sheets. Google sheets currently allow Dashboard users to track specific teachers by grouping them based on their email addresses and location.

B. TABLEAU WORKBOOKS

Once the data tables are ready, they can be connected with **Tableau Desktop -** the software required to create **dashboards.** It is a licensed software, which is free for up to a year for university students.

Tableau creates a local datasource for itself called an **Extract**, which it uses to prepare **workbooks**. A workbook contains **sheets**, several of which can then be connected to prepare a **Dashboard.**

These workbooks can then be hosted on the **Tableau server**.

C. TABLEAU SERVER

Tableau server is a platform where workbooks can be uploaded and interacted with. WPI **Institutional Research team** maintains a tableau server at tableau.wpi.edu, which is also a licensed platform. Also, the server sets up a scheduled refresh of the data extract connected to the workbooks.

People in the ASSISTments team need to have "**@wpi.edu**" email addresses and the WPI VPN to be able to request Tableau server access.

Once that is done, the workbooks can be accessed by a user friendly web based UI. These workbooks are set to refresh every day, providing the latest data in a user friendly dashboard.

3. DATA PREPARATION

There were two priorities in the Data Preparation stage:

A. Understand end user requirements.
B. Prepare persistent data tables for Tableau to connect with.

A. Understand end user requirements : Up to date Analysis and Reporting

Initially, the project started with counting the teachers signing up and tracking the Assignments made by teachers. The main functionalities required were analysis and reporting of the data. Also additional features were added based on user feedback over time as they used the dashboard. For example see Figure 3 and 4 below.

From: Tignor, DonnaLee
Sent: Tuesday, November 10, 2020 8:54 AM
To: Jakhmola, Rahul
Subject: Re: [EXT] Re: Tableau Data

Hi Rahul
It worked! I can now access this new report through the link that you provided. However, this new version doesn't provide important info needed for evaluating the mentee's use of ASSISTments as compared to the old version in individual reports. For example, the old version provided me with 1) individual assignment dates for each assignment, 2) how many times each individual report was accessed by the teacher, 3) individual assignments number of problems completed as well as 4) individual assignment completion rates. Is there any way this info can be accessed for individual reports without having to click in and out of each assignment? It is an important component in my evaluation. In the current report, I can't tell at a glance, whether or not a mentee has been recently assigning, checking reports, and if students are completing at a proficient rate since the last time we met. Thanks for your help.

Kind regards,
DonnaLee

Fig 3 : User feedback example -
DonnaLee Tignor, Coordinator at Montachusett Regional Vocational Technical School District

Cristina 4:32 PM
@Rahul Jakhmola I really like your new open response dashboard. Can you allow me to put in a Problem ID and get to that spot. So I can ask myself Are there any examples of OR answers for this problem. Also can it use the PR ID not the number. Also how about the text answers from students could you load both and then have a filter to filter out text and just look at uploaded.

One more thing if I sort by Correct can they stay clumped by problem ID?

Fig 4 : User feedback example -
Cristina Heffernan, Executive Director of the ASSISTments Foundation

As the Product, Marketing and Communications team got involved over time, more dashboards and data sources were added in ongoing iterations. More features like curriculum used, reports checked, student completion rates and links to assignment

reports were needed to be added. Also, the users needed to download a list of relevant teachers along with their summary statistics.
Currently, there are three separate **Folders** for our **33 tableau dashboards** on the server - **Archive (23 Dashboards), Live(3 Dashboards) and Test (7 Dashboards).**

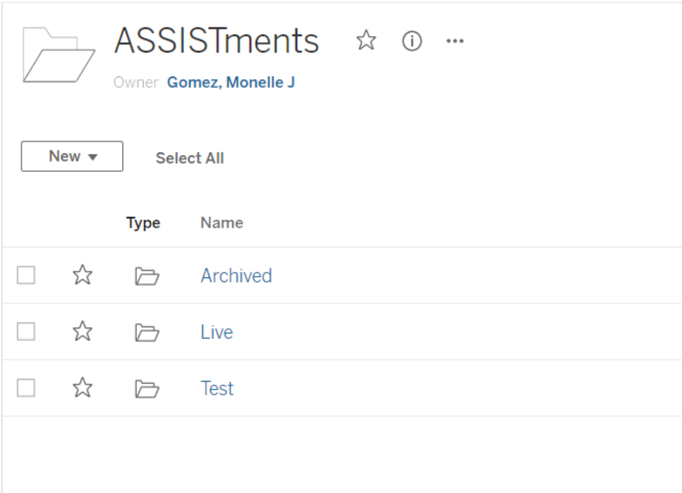


Fig 5 : ASSISTments home folder on tableau.wpi.edu

Dashboards with features that were no longer needed were archived. New features being tested out are kept on beta versions of the dashboard. Everything else stays in the Live folder.

B. Preparing persistent Data Tables for Tableau

There were two main challenges that were addressed in this section.

1. Combining Data from different schema : FORIEGN DATA WRAPPER

Tableau provides excellent filters and data filtering when connected to up to two data tables. So there was a requirement to do data preparation on the server itself, and have just one PostgreSQL table for Tableau to connect with.

To address this issue, there was a need to combine data from three different schema - ASSISTments 1.0, ASSISTments 2.0 and ASSISTments 2.0 logs.

This was done by using the [Foriegn Data Wrappers] in PostgreSQL, which allow the user to query multiple databases from the same schema.

Also, there were some tables that were most important for answering the user questions. Schema diagrams for joining those tables are shared in the Appendix.

2. **Data Persistence after daily refresh : SQL Views**

   Also, once the tables were prepared, they needed to persist on the server after the server goes through a daily data refresh. This was done by creating **SQL scripts** that create **views** on the server.

   A **view** is essentially a query which can be called like a table. Whenever Tableau calls to a view, it runs a query and makes Tableau believe that it is looking at a table. We currently have two scripts running for ASSISTments 1.0 and ASSISTments 2.0.

   Tableau connects to just ASSISTments 2.0 for most dashboards, where it calls to views to get the data.

4. USER FRIENDLY DASHBOARD DESIGN

The aim was to create intuitive visualizations that allow users to subset, pivot and download their data. This was done through creating Key Indicators, and then using standard interactivities on the dashboards.

Key indicators were created for Assignments, which could then roll up to teacher, domain and state levels.

A. Key Indicators

For every assignment that was created, the following numbers were calculated:

- **Class Size**
- **Started By**
- **Completed By**
- **Completion Rate (%)**
- **Report Checked by Teacher (Yes/No)**
- **Curriculum Used**
- **First Response by student Timestamp**
- **First Report view by teacher Timestamp**
- **Is a LOOP (T/F)**

  What is a loop?

  A loop is defined as a teacher viewing an assignment report after the first student has responded to it. This definition comes from 4 simple steps of ASSISTments.

  This indicator allows us to track how many teachers actually look at an assignment report and gain insight from it after a student has responded.

B. Interactivities on Dashboards

Tableau dashboards come with standard interactivities like drop downs, clickable visualizations and filters that can be customized as per our data.

These filters have turned out to be very user friendly, and are utilized by end users with very little support.

Some examples of interactivities are as follows.

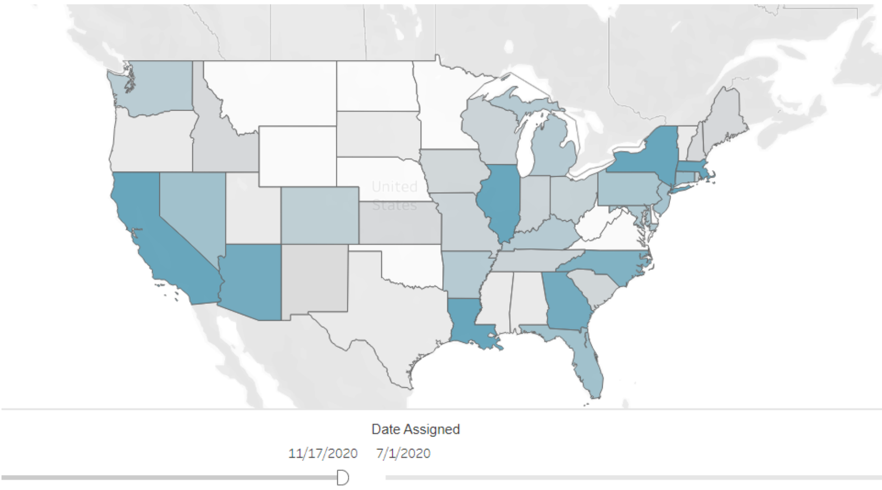


Fig 6 : Map Interactivity - You can click on states to filter data

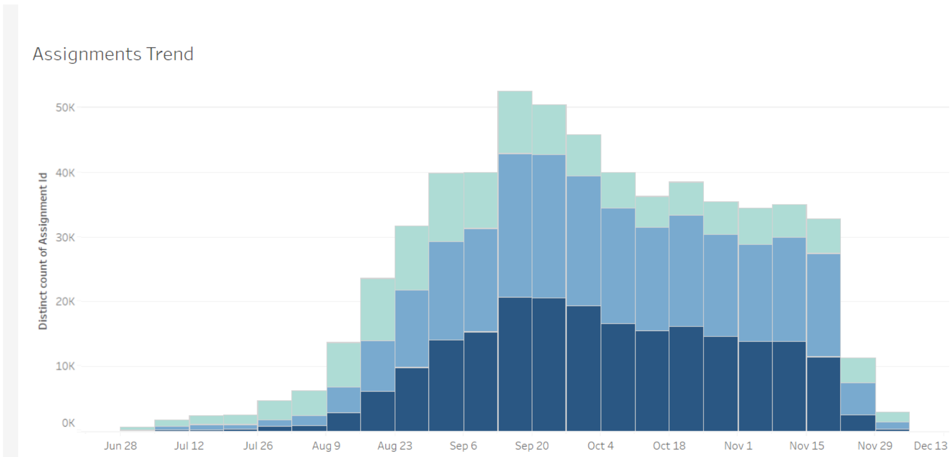


Fig 7 : Chart Interactivity - You can click on dates to filter data by week

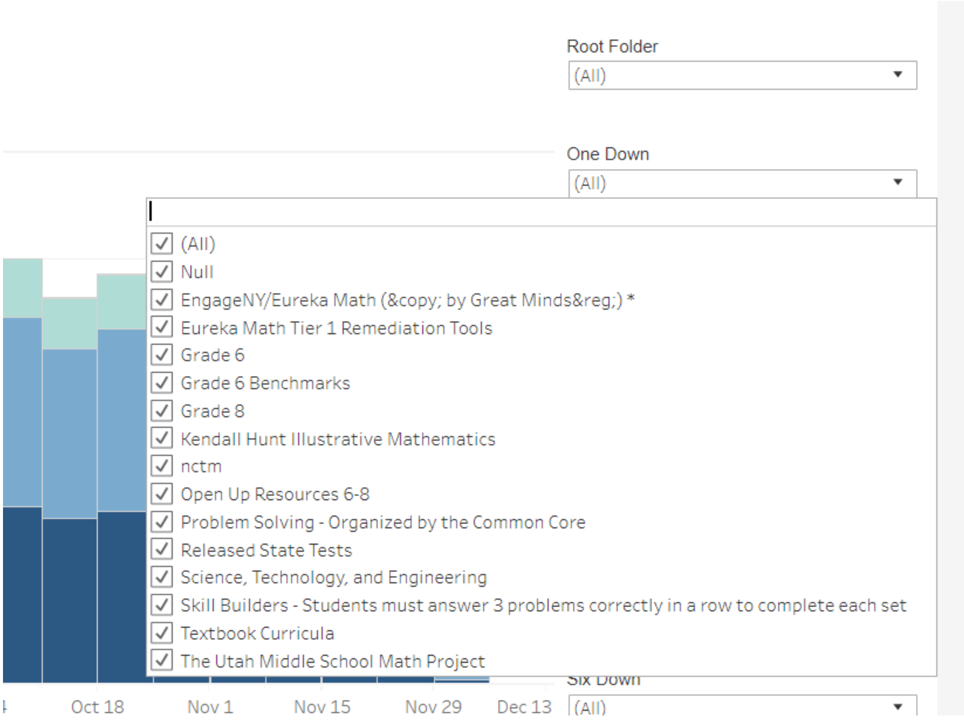


Fig 8 : Drop Down Interactivity - You choose different curriculum to view

5. CREATING FUNCTIONAL PROCESS

We have a four step process to getting a dashboard up and running from concept to production. A training was also delivered to stakeholders who would want their own dashboards.

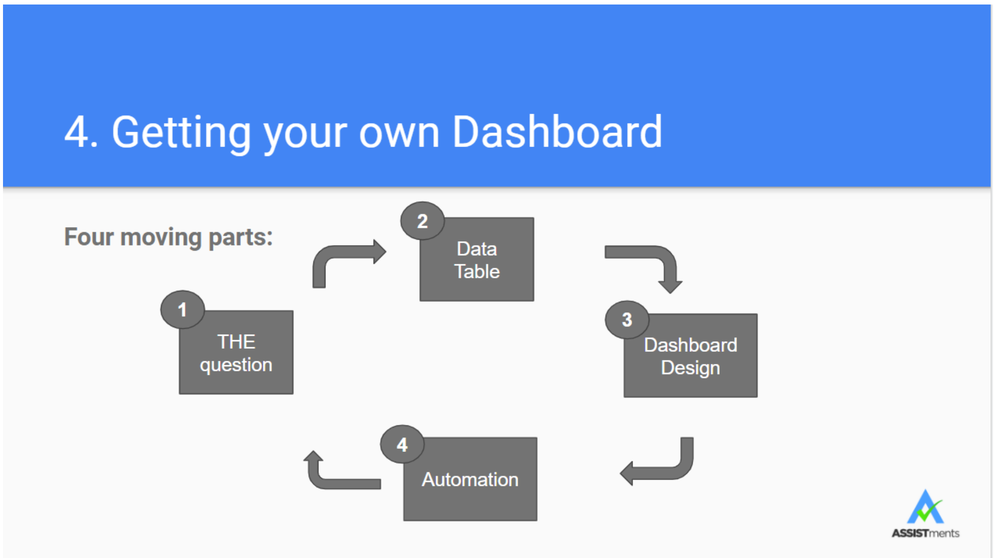


Fig 9 : Four step process for getting personalized dashboards.

The following guidelines are from the training conducted.

Part 1 : The Question

What does the data say about XYZ? Any curiosity first needs a person asking the question.

Getting clarity about the question you need answering would help in the data preparation and dashboard development.

Part 2 : Data Preparation and Automation

Once the question is asked, we can look for the data in our database. (And possibly google analytics and other sources.)

For this our tools are SQL, R or Python for aggregations on the server and maybe some API calls that allow us to pull the required data and then dive into it.

Also, all these steps need to be automated to provide regular updates over time.

Part 3 : Dashboard Development

Once the prepared data is accessible, Tableau Dashboard development can start. The prepared data source connects to a dashboard, where reports, visualizations and pivots are prepared for a visual analysis.

This requires a full version of Tableau Desktop which incurs a licensing fee.

It is free for a year for students. WPI also has a Tableau team and has purchased Tableau Desktop licenses which can be requested if required.

Part 4 : Automation

When the dashboards developed are ready to go live, they need to be hosted on a Tableau Server. WPIs Tableau team maintains a server which is protected by a VPN. Any access required over here needs to be requested from them. You would need a wpi.edu account to log in through the single sign on.

Once you have a wpi.edu id, the server allows for multiple roles - if you consume data from dashboards you need to request a viewer access, and will be added to a viewer group. It’s best to route this through Cristina and Rahul.

The dashboard developer has write access and can control who can view the dashboards. The role also needs to be requested, besides having Tableau Desktop.

The server is also responsible for regular refreshes of the dashboard from the automated data sources prepared initially.

5. SUCCESSES, CONCLUSION AND SCOPE

This project has been successful in getting adopted for day to day use by ASSISTments staff.

Currently, two most used dashboards are helping Comms, Product and Teacher Experience teams make decisions daily.

The mailing campaign dashboard provides an up to date list of emails to Kristyn Manoukian and lets her keep track of who has received email campaigns.

The Assignment Activity dashboard is the most used dashboard, and is used by the ASSISTments team to get the required information they need. This dashboard has had multiple iterations till date and has been viewed upwards of 2000 times.

Current scope also involves setting up standard data tables for other students and employees to create their own Dashboards. This project is already providing support to other projects for the IT team.

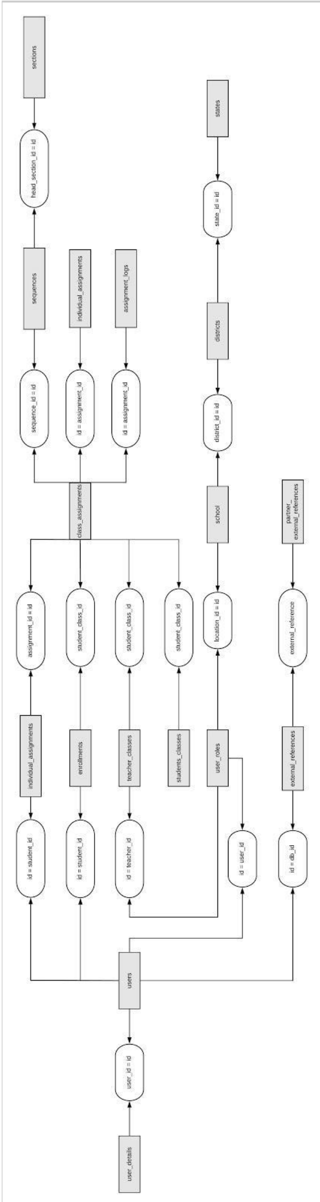
There are more dashboards and ideas in the pipeline which will enable more real time analytics, and help us track insights and teachers in a more actionable way.

6. ACKNOWLEDGEMENT

This project has been highly dependent on questions and feedback being driven by the end users involved.

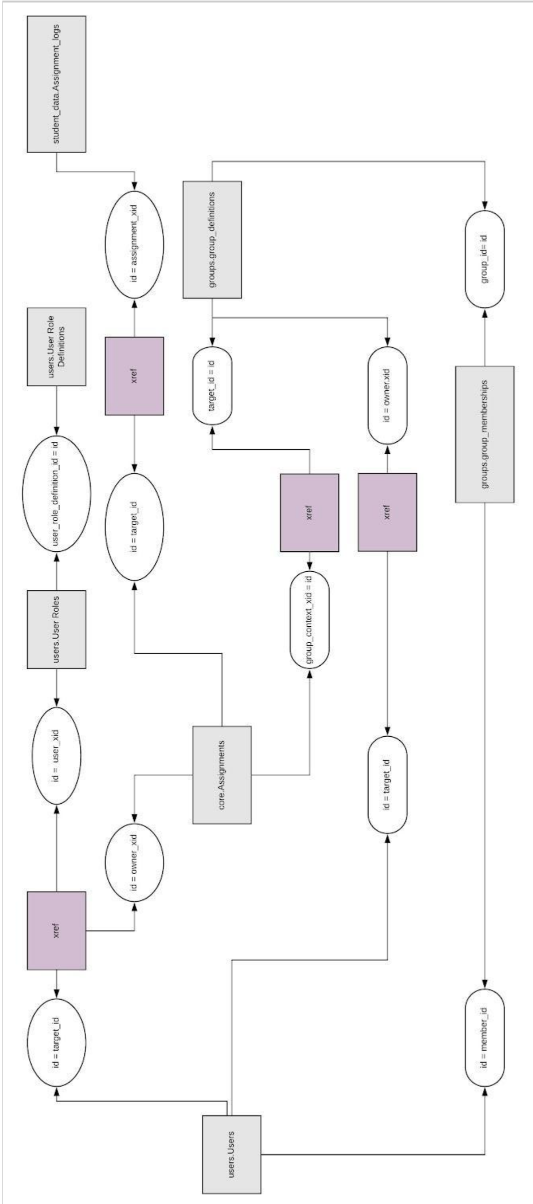
Appendix 1 for reference : Schema for 1.0 Tables

Schemas - 1.0.png



Appendix 2 for reference :Schema for 2.0 Tables

[Schemas - 2.0.png]



Project 2 : A/B Testing

Influencing Individual Teacher Behavior

0. Overview

It is important for teachers to monitor the progress of their students in classrooms. In the last year, due to COVID-19, a large number of teachers resorted to using computer based learning systems. In computer based learning systems, student progress is monitored via data reports.

However some teachers need help navigating these computer based learning platforms to access student reports and uphold the same student-teacher interactions they would normally have in classroom settings.

A sizable number of assignment reports available were not viewed by the teachers in the previous school year (2019-2020). Without access to student responses, teachers are unable to incorporate valuable information in instructional strategies.

36%
(4,930/13,606)
Teachers never checked their assignment reports.

41%
(141,102/343,997)
Assignments had student responses but the report was never checked.

To address the research questions, a study was conducted within the ASSISTments online learning platform that compares 3 experimental conditions (2 treatment conditions and1 control condition).

The two treatment conditions each compare a different format of email prompt encouraging teachers to check on reports of their students' work. These Differing prompts are illustrated in Figure 1. The first condition, referred to as the "no call to action" condition and shown on the right in that figure, simply notifies teach-ers that reports of student work are available to be viewed within the system. The second treatment condition, referred to as the "call to action" condition and shown on the left in the figure, similarly notified teachers of available reports,but also provided a direct link to view such reports.

These Treatment conditions were compared to a control condition where no such email prompt was sent to teachers.The intuition here was that the "no call to action" conditions would allow us to compare what the effect is of sending email prompts on teacher engagement.

Given that it is important for teachers to monitor student progress, we hypothesized that these prompts would help remind teachers that reports are available to them within the system. The “call to action”condition then allows us to further observe whether providing a direct link to reports, helping teachers who may not know how to navigate to their students’ work, leads to a greater effect on teacher engagement over the other conditions

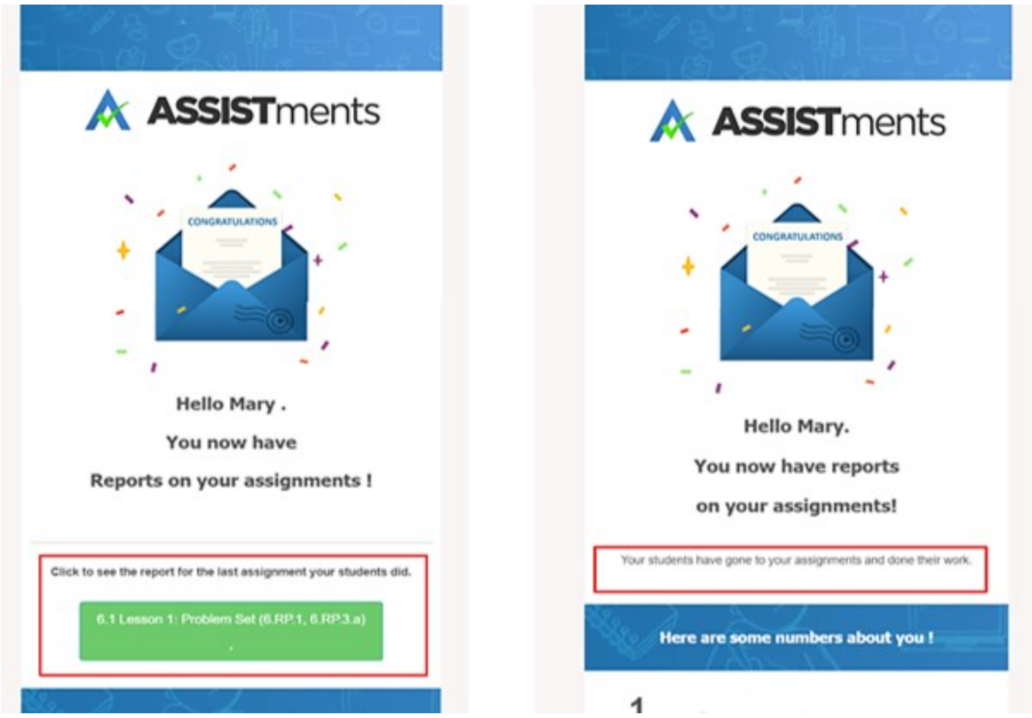


Figure 1: Examples of “Call to action” (left) and “No call to action” (right) email prompts.

The following research paper came out of this project. I intend to submit it to upcoming conferences around learning analytics.

Do Email Prompts Increase Teacher Engagement Within Computer-based Learning Systems?

Author 1
Affiliation 1
email 1

Author 2
Affiliation 2
email 2

Author 3
Affiliation 3
email 3

ABSTRACT

It is important for teachers to monitor the progress of their students in classrooms. In the last year, due to COVID-19, a large number of teachers resorted to using computer based learning systems. In computer based learning systems, student progress is monitored via data reports. However some teachers need help navigating these computer based learning platforms to access student reports and uphold the same student-teacher interactions they would normally have in classroom settings. We ran a study to compare different methods of prompting teachers to view reports of their students' work. We sent emails and compared two types of email prompts to a control group. We compare the effect of this email prompt across several teacher activity measures. We find no measurable improvement in teacher report checking, number of assignments, or student engagement when controlling for prior interaction with the system.This paper presents a null finding, but offers an exploration into methods designed to increase teacher involvement in their students' learning.

Keywords

ASSISTments, Marketing Campaign,Education Data Mining

1. INTRODUCTION

Numerous studies have found correlations between teachers' monitoring of student progress and later academic achievement [6, 10, 12, 13]. When teachers are provided with information on their students, they may be able to better direct their attention to subject areas and students in need of greater attention or remedial instruction (c.f. [9]); these and other works have led to the development of teacher augmentation tools and frameworks designed to equip teachers with better supports in monitoring student progress within computer-based systems [3]. It has been found, perhaps unsurprisingly, that providing teachers with data can improve their ability to anticipate areas where their students are likely to experience difficulty [8], allowing them to incorporate this knowledge into their instructional practices.

However, these supports can only be effective when teachers are able to attend to and interpret such information; teachers who do not view reports within computer-based systems are unable to incorporate that valuable information into their instructional strategies. In most computer-based learning platforms, teachers access their students' data through generated reports and dashboards [9, 1, 4]. Whether static, or dynamically reporting in real time [9, 6], this information is meant to summarize various aspects of student performance and behavior in a manner that helps teachers take action. In some cases, however, the amount of information presented in these reports and dashboards may be overwhelming to a teachers' cognitive load [5]. If reports are difficult to interpret, a teacher may fail to see value in the information.

It is important for developers of learning platforms to provide teachers with the right set of supports to promote positive teacher-student engagement in differing learning environments; this is particularly highlighted during the period of remote learning in response to the COVID-19 pandemic. It is important for the field to explore methods of guiding teachers to those practices that have been found to lead to positive impacts on learning. In many commercial and marketing contexts, email prompts have often been utilized to increase user engagement [15, 7, 11]. In education contexts, email has also been utilized to similarly engage users within massive open online courses (MOOCs; [14]). While the effectiveness of such prompts have exhibited varying, though often positive, success [2], it is the goal of this work to explore the use of such prompts in encouraging teachers to attend to their students' data in a K-12 remote classroom setting.

In this paper, we present on a randomized controlled trial that explores the effectiveness of email prompts on two outcomes of teacher engagement within a computer-based learning platform. Within this, we seek to address the following research questions:

1. Does the sending of email prompts increase teacher engagement as measured through assigning and report-checking activity?
2. Does the inclusion of a direct link to a report within email prompts increase teacher engagement?

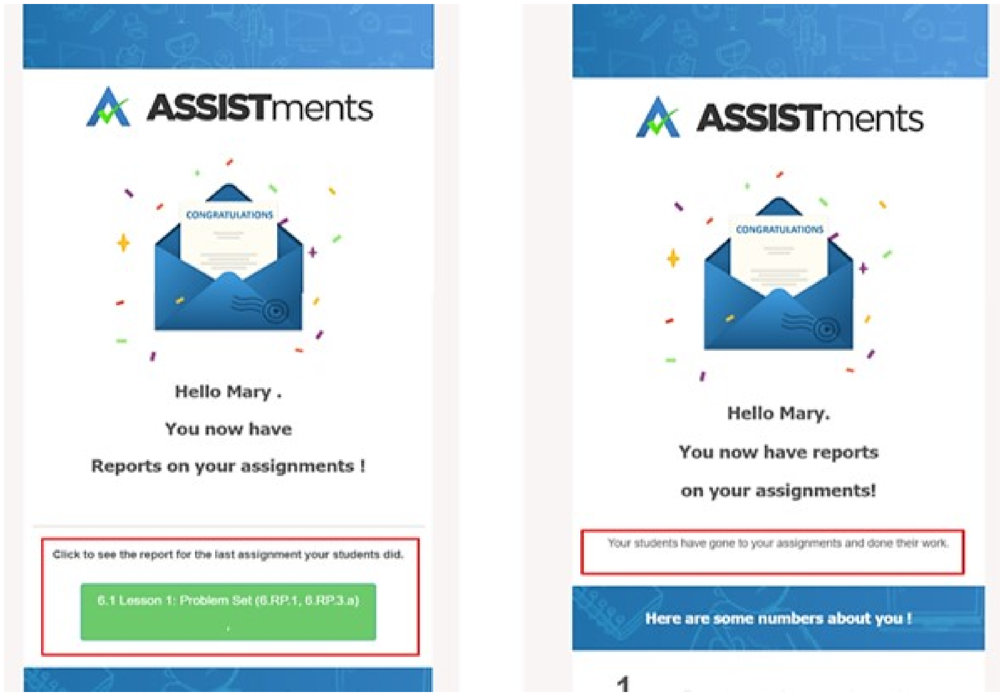


Figure 1: Examples of "Call to action" (left) and "No call to action" (right) email prompts.

2. EXPERIMENTAL DESIGN

To address our research questions, we conduct a study run within the ASSISTments online learning platform that compares 3 experimental conditions (2 treatment conditions and 1 control condition). The two treatment conditions each compare a different format of email prompt encouraging teachers to check on reports of their students' work. These differing prompts are illustrated in Figure 1. The first condition, referred to as the "no call to action" condition and shown on the right in that figure, simply notifies teachers that reports of student work are available to be viewed within the system. The second treatment condition, referred to as the "call to action" condition and shown on the left in the figure, similarly notified teachers of available reports, but also provided a direct link to view such reports. These treatment conditions were compared to a control condition where no such email prompt was sent to teachers.

The intuition here was that the "no call to action" condition would allow us to compare what the effect is of sending email prompts on teacher engagement. Given that it is important for teachers to monitor student progress, we hypothesized that these prompts would help remind teachers that reports are available to them within the system. The "call to action" condition then allows us to further observe whether providing a direct link to reports, helping teachers who may not know how to navigate to their students' work, leads to any greater effect on teacher engagement over the other conditions.

2.1 Study Context and Participants

ASSISTments is a free computer-based learning platform, focused primarily on middle school mathematics, that allows teachers to build and assign homework and classwork online. The progress of students on these assignments are available for teachers in assignment reports. Teachers can monitor these reports to observe their students' performance on the assignments and incorporate that knowledge into their instruction. This cycle of assigning content, having students complete that assigned material, and then viewing reports of the student work, referred to as a full "loop," is important as it is an indicator of teacher engagement (not only with the system, but, more importantly, with their students' learning).

However, as found in an initial exploratory analysis, a sizeable number of assignment reports available are not viewed by the teachers. Nearly **36%** of teachers (4930 / 13606) had never checked their assignment reports. About **39%** of assignments (133,654 / 343,997) had the teachers viewing progress reports. **41%** assignments (141,102 / 343,997) had student responses on them but the teachers had not looked at reports.

Given this large number of teachers who seemingly had not been viewing reports, we target these teachers for our study. We sampled 618 teachers who had made accounts in the system between July 1, 2020 and September 21, 2021, made at least one assignment, but completed no loops (again, referring to the cycle of assigning, students working, and report viewing). The objective was to see if sending them emails with directions to view assignment reports could lead them to engage more with student reports.

2.1.1 *Randomization, Treatment, and Outcomes*

In order to ensure a roughly-even distribution of teachers within and across conditions, a stratified randomization was applied. We observed four teacher attributes which we hypothesized to likely relate to teacher engagement in the future. These attributes are:

- **Age** (How long ago they had made accounts)
- **Assignments** (How many assignments they had made)
- **Classes** (Number of classes they had created)
- **Students** (Number of students they had in classes)

It is important to note that since we observed newly-created accounts within a limited time span, the age of the account is likely capturing which teachers started during summer months versus those who started after the beginning of the school year in September, as opposed to, for example, teacher experience.

Stratification was applied using k-means clustering across four variables, resulting in 14 different clusters. From these clusters, we applied stratified sampling into 3 groups representing assignment to each of the 3 experimental conditions. This ensures that each condition contains representatives from all strata.

The teachers in the two test groups were sent just a single email on September 21 6:30 pm EST. Following the first half of the academic year, observations of teacher engagement were then made on February 8, 2021. Specifically, we observe both the number of complete assignment/report loops for each teacher, as well as the number of assignments given by each teacher.

3. DATASET

The variables collected for each teacher within our study are listed and described in Table 1. Within our analyses, described in the next section, each teacher is represented by 5 variables, including the 3-value categorical variable of assigned condition.

Of the 618 teachers who participated in the study, it was found that only 240 created any new assignments within ASSISTments between the dates observed for the study (September 2020 - February 2021). Among these 240 teachers, only 127 completed at least one loop. The breakdown of these statistics are reported in Table 2.

4. APPROACH TO ANALYSIS

A chi-square test of independence showed that there was no statistically reliable difference across the three groups in regard to the number of teachers who created at least one new assignment: X2 (2, N = 618) = 0.8583, p = 0.6511. As such, we focus our analysis only on the 240 teachers who remained active within ASSISTments (by created at least one assignment after being assigned to condition).

In order to address our two research questions, we conduct two regression analyses observing the relationship between the assigned condition with each of the two dependent measures identified in Table 1, while controlling for the other covariates.

Variable	Description
Independent Variables	
Account Age	Age of the teacher's account (in days) before condition assignment
Prior Assignments	Number of assignments made up to the time of condition assignment
Number of Classes	Number of classes created by the teacher at the time of condition assignment
Number of Students	Number of students enrolled across all of the teacher's classes
Assigned Condition	Experimental condition to which the teacher was assigned
Dependent Variables	
Number of New Assignments	Number of new assignments created after condition assignment
Number of Completed Loops	Number of assignment/report loops completed by the teacher after condition assignment

Table 1: Description of variables observed in the collected experimental data.

Group	Total Teachers	Assignment Creators	Loop Creators
No email (Control)	206	84	39
No Call to Action	206	81	46
Call to Action	206	75	42

Table 2: Counts of teachers in the three groups.

Due to the distribution of many of these variables, including the dependent variables, several transforms were applied to convert these into approximate-normal distributions. Specifically, a log transform is applied to each dependent variable (or $log(loops + 1)$ in the case of the number of loops due to some teachers completing 0 loops), while either log or square root transforms are applied to other covariates; these are specified later in Table 3.

5. RESULTS

The results of both regression analyses, observing each dependent measure of teacher engagement, is reported in Table 3. As can be seen in that table, neither condition exhibits statistically reliable effects in regard to either outcome. While the coefficient of these do lean in favor of the email prompts exhibiting a positive effect on the number of loops completed (with the call to action exhibiting a slightly higher coefficient), this difference is not statistically reliable. This effect appears to be even smaller in the case of the number of new assignments given.

6. DISCUSSION AND FUTURE WORK

The conducted study represents, in this context, a null finding. While this does not suggest that email prompts have no effect, it does mean that the effect is too small for us to measure given the sample size of teachers. While it is often

Dependent Variable:	Assignment/Report Loops			Log Number of Assignments		
Variable	B	Std. Error	p-value	B	Std. Error	p-value
Intercept	1.292	0.331	<0.001***	2.550	0.354	<0.001***
Email with No Call to Action (Treatment 1)	0.201	0.226	0.376	0.003	0.242	0.991
Email with Call to Action (Treatment 2)	0.243	0.227	0.285	-0.059	0.243	0.807
Log of Prior Teacher Assignments	0.062	0.111	0.575	0.580	0.119	<0.001***
Square Root of Age of Teacher Account	-0.100	0.057	0.079	-0.046	0.060	0.447
Log of Number of Students	0.112	0.083	0.179	-0.190	0.089	0.034*

Table 3: Regression analysis results observing the log of assignment/report loops completed by teachers (left) and the log number of assignments created (right) as dependent variables.

difficult to make any strong conclusions from a null result, we believe that this result is particularly interesting given the context. It is surprising that raising awareness of available reports would motivate teachers to utilize these within the learning system; we posit that it seems questionable to think that a teacher would disagree with the idea of reports of student work being unhelpful.

There could be several explanations that might explain the lack of effect here despite similar email campaigns having shown promise in other contexts. First, it could be that many of these teachers are overwhelmed with the amount of information made available to them; teachers may be recording or monitoring student progress in different ways external to the observed learning platform. Alternatively, and very likely, teachers may not have viewed the email to begin with; we did not incorporate any tracking in this study to observe the number of teachers who ultimately opened the email.

It is important to explore these aspects through future research, particularly to understand whether teachers are becoming overwhelmed while engaging with computer-based learning platforms. If teachers are not finding value in the developed tools and supports made available, such information could help inform improvements to such systems.

7. REFERENCES

[1] V. Aleven, F. Xhakaj, K. Holstein, and B. M. McLaren. Developing a teacher dashboard for use with intelligent tutoring systems. *technology*, 34:44–50, 2010.

[2] G. Alkhaldi, F. L. Hamilton, R. Lau, R. Webster, S. Michie, and E. Murray. The effectiveness of prompts to promote engagement with digital interventions: a systematic review. *Journal of medical Internet research*, 18(1):e6, 2016.

[3] P. An, K. Holstein, B. d'Anjou, B. Eggen, and S. Bakker. The ta framework: Designing real-time teaching augmentation for k-12 classrooms. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–17, 2020.

[4] R. Bodily, J. Kay, V. Aleven, I. Jivet, D. Davis, F. Xhakaj, and K. Verbert. Open learner models and learning analytics dashboards: a systematic review. In *Proceedings of the 8th international conference on learning analytics and knowledge*, pages 41–50, 2018.

[5] D. F. Feldon. Cognitive load and classroom teaching: The double-edged sword of automaticity. *Educational Psychologist*, 42(3):123–137, 2007.

[6] M. Feng and N. T. Heffernan. Towards live informing and automatic analyzing of student learning: Reporting in assistment system. *Journal of Interactive Learning Research*, 18(2):207–230, 2007.

[7] C. S. Garland, K. C. Yang, and Y. Kang. The effectiveness of permission-based email marketing for non-profit organizations: A case study of national public radio (ktep-fm) in el paso, texas. *Journal of Creative Communications*, 11(1):17–26, 2016.

[8] H. C. Hill, D. L. Ball, and S. G. Schilling. Unpacking pedagogical content knowledge: Conceptualizing and measuring teachers' topic-specific knowledge of students. *Journal for research in mathematics education*, 39(4):372–400, 2008.

[9] K. Holstein, B. M. McLaren, and V. Aleven. Student learning benefits of a mixed-reality teacher awareness tool in ai-enhanced classrooms. In *International conference on artificial intelligence in education*, pages 154–168. Springer, 2018.

[10] B. D. Nye. Barriers to its adoption: A systematic mapping study. In *International Conference on Intelligent Tutoring Systems*, pages 583–590. Springer, 2014.

[11] R. Rettie, U. Grandcolas, and V. Payne. Email marketing: Permission to pester. In *Academy of Marketing (AM) Annual Conference*, pages 2–5. Citeseer, 2002.

[12] S. Ritter, M. Yudelson, S. E. Fancsali, and S. R. Berman. How mastery learning works at scale. In *Proceedings of the Third (2016) ACM Conference on Learning@ Scale*, pages 71–79, 2016.

[13] J. W. Schofield. *Computers and classroom culture.* Cambridge University Press, 1995.

[14] J. J. Williams, A. N. Rafferty, S. Maldonado, A. Ang, D. Tingley, and J. Kim. Mooclets: a framework for dynamic experimentation and personalization. In *Proceedings of the Fourth (2017) ACM Conference on Learning@ Scale*, pages 287–290, 2017.

[15] A. Yasmin, S. Tasneem, K. Fatema, et al. Effectiveness of digital marketing in the challenging age: An empirical study. *International Journal of Management Science and Business Administration*, 1(5):69–80, 2015.

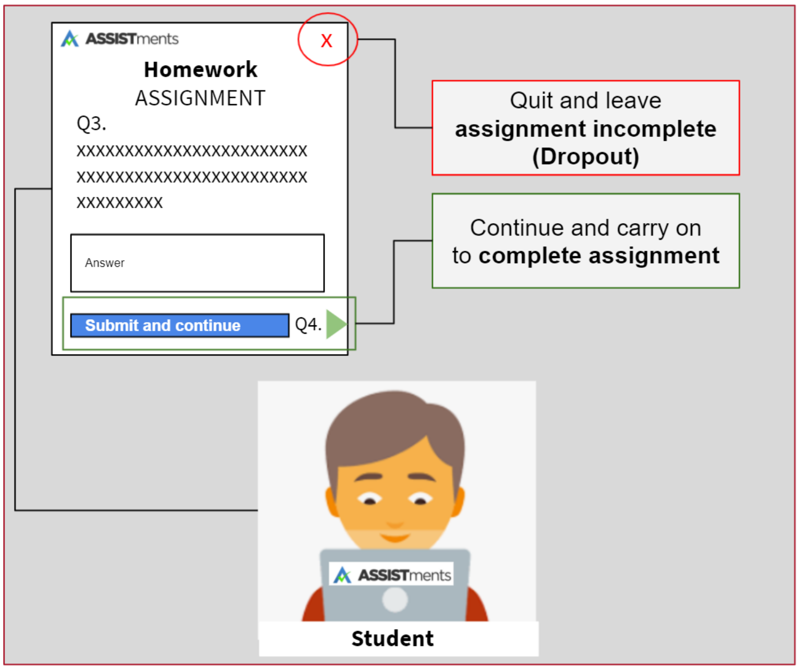
Project 3 : Predictive Analytics

Influencing Individual Student Behavior

1. Introduction

After exploring missed learning opportunities by teachers in the last project, this project aimed to look at missed learning opportunities by students. The analysis was done on data from the ASSISTments platform from the 2018-2019 school year.

In Massachusetts, a total of 7,918 students did 203,268 homework assignments. However,over 8% of the assignments were left incomplete, as students left - or dropped out - in the middle of the assignments. This becomes over 16,000 missed learning opportunities in an assignment.

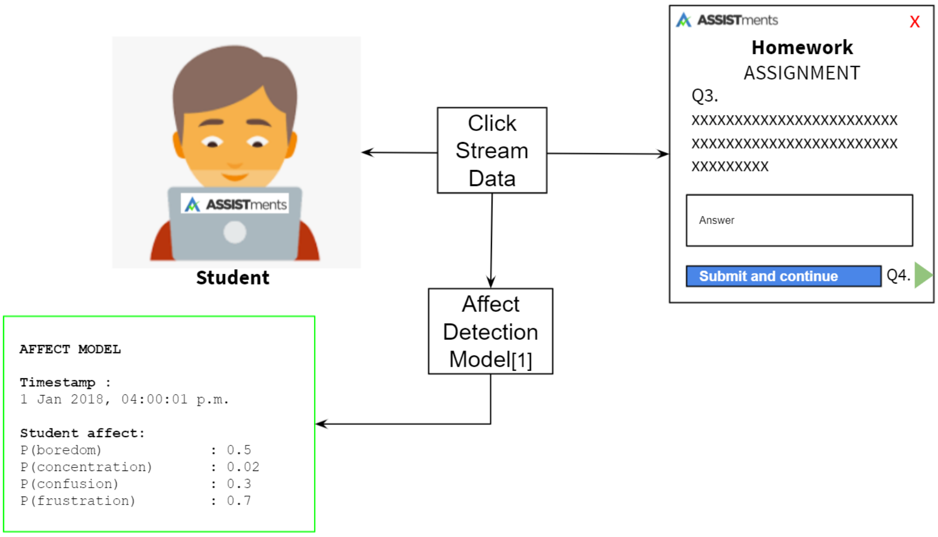


Could this dropout be attributed to student emotion? In this project we explore how much student emotion during assignments is associated with dropout in math assignments. We also aim to tease apart whether the dropout in an assignment is based on students' propensities to certain emotions (i.e. student emotional affective trait) or because of interactions with certain assignments (i.e. student emotional affective state while solving that assignment).

Students experience emotions like boredom, confusion or frustration while doing their homework and which can even make them stop completely in the middle of a homework assignment altogether. High math frustration has already been shown to be associated with lower STEM outcomes and reduced college attendance. (high math frustration)

2. Tracking Student Emotion : Student Affect

Student emotion data, or student affect, can be estimated using state of the art deep learning based affect detection models (Botelho et al). The models can help predict student emotion probabilities (affect) by just observing the clickstream data of how students interact with online learning platforms, in this case ASSISTments.



The affective states observed include boredom, concentration, confusion and frustration. As multiple students interact with multiple assignments, their propensity for a particular affect can be calculated. Same is the case with assignments. IRT models like the Rasch model (reference) can help tease apart student level and assignment level affect propensities.

What this means is that based on past interactions, the propensity of students towards boredom or assignments inducing boredom can be calculated as a z-score. This is valid for other affects like confusion, concentration and boredom as well.

3. Data Overview

A logistic model is developed to find variables that are associated with student dropout in the assignments.

- Every row of data is an observation of a student dropping out of an assignment.
- For every time a student drops out, their propensities for affect and dropping out are recorded. Similarly, the propensity of that assignment for the affect is recorded.
- Another factor taken into consideration while modeling was the prior dropout rate of students. If a student has dropped out on two assignments before the current one, the prior dropout would be 2.

Logistic regression was run on these 16,000+ rows of data.

LOGISTIC MODEL

```
Dropout ~
      Prior_dropout +

      student_concentration_propensity +
      student_boredom_propensity +
      student_confusion_propensity +
      student_frustration_propensity +

      question_concentration_propensity +
      question_boredom_propensity +
      question_confusion_propensity +
      question_frustration_propensity
```

4. Analysis

Coefficients:

	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	-2.8776825	0.0103125	-279.047	< 2e-16	***
prior_dropout	3.9012357	0.0456599	85.441	< 2e-16	***
student_concentration_propensity	-0.0029178	0.0022828	-1.278	0.20118	
student_boredom_propensity	0.0006981	0.0022136	0.315	0.75249	
student_confusion_propensity	-0.0031933	0.0023846	-1.339	0.18052	
student_frustration_propensity	-0.0076202	0.0025617	-2.975	0.00293	**
question_concentration_propensity	-0.0052919	0.0017588	-3.009	0.00262	**
question_boredom_propensity	0.0077022	0.0017155	4.490	7.13e-06	***
question_confusion_propensity	-0.0016415	0.0015136	-1.085	0.27814	
question_frustration_propensity	-0.0016226	0.0014365	-1.130	0.25868	

Prior dropout is the strongest predictor of future dropout in assignments.

Affect propensity values for students and assignments appear to be significant, but show small coefficients.

- Lower student frustration is significantly associated with dropping out.
- Lower question concentration propensity and higher question boredom propensities are associated with dropouts.

5. Conclusion and Future Scope

Paying timely attention to students who drop out in online learning courses could help understand what kind of interventions are needed to prevent further dropout.

Further analysis of student dropout rates across multiple socio-economic conditions can be done in the future for these students.