Integrated Closed-loop Learning Analytics Scheme in a First Year Experience Course

Munira Syed
Notre Dame, IN, USA
msyed2@nd.edu

Trunojoyo Anggara
University of Notre Dame
Notre Dame, IN, USA
tanggara@nd.edu

Alison Lanski
University of Notre Dame
Notre Dame, IN, USA
alanski@nd.edu

Xiaojing Duan
University of Notre Dame
Notre Dame, IN, USA
xduan@nd.edu

G. Alex Ambrose
University of Notre Dame
Notre Dame, IN, USA
gambrose@nd.edu

Nitesh V. Chawla
University of Notre Dame
Notre Dame, IN, USA
nchawla@nd.edu

ABSTRACT

Identifying non-thriving students and intervening to boost them are two processes that recent literature suggests should be more tightly integrated. We perform this integration over six semesters in a First Year Experience (FYE) course with the aim of boosting student success, by using an integrated closed-loop learning analytics scheme that consists of multiple steps broken into three main phases, as follows: Architecting for Collection (steps: design, build, capture), Analyzing for Action (steps: identify, notify, boost), and Assessing for Improvement (steps: evaluate, report). We close the loop by allowing later steps to inform earlier ones in real-time during a semester and iteratively year to year, thereby improving the course from data-driven insights. This process depends on the purposeful design of an integrated learning environment that facilitates data collection, storage, and analysis. Methods for evaluating the effectiveness of our analytics-based student interventions show that our criterion for identifying non-thriving students was satisfactory and that non-thriving students demonstrated more substantial changes from mid-term to final course grades than already-thriving students. Lastly, we make a case for using early performance in the FYE as an indicator of overall performance and retention of first-year students.

KEYWORDS

intervention, at-risk students, learning analytics, first year seminars, first year experience, advising

ACM Reference Format:
Munira Syed, Trunojoyo Anggara, Alison Lanski, Xiaojing Duan, G. Alex Ambrose, and Nitesh V. Chawla. 2019. Integrated Closed-loop Learning Analytics Scheme in a First Year Experience Course. In *The 9th International Learning Analytics & Knowledge Conference (LAK19), March 4–8, 2019, Tempe, AZ, USA.* ACM, New York, NY, USA, Article 4, 10 pages. https://doi.org/10.1145/3303772.3303803

LAK19, March 4–8, 2019, Tempe, AZ, USA

ACM ISBN 978-1-4503-6256-6/19/03…$15.00
https://doi.org/10.1145/3303772.3303803

1 INTRODUCTION

Identifying at-risk students is an established area of research in learning analytics [7, 27, 29, 40], whereas an emerging area explores the design of learning analytics interventions [39]. There is not much research, however, that attempts to combine the two and close the learning analytics loop. Furthermore, to the best of our knowledge, existing studies do not examine the evolution and evaluation of intervention mechanisms over the years when a course is offered multiple times. A possible reason for the lack of such studies is the problem of designing an infrastructure that will make this analysis possible. In this paper, we aim to show how the combination of learning data, platform design infrastructure, identification of non-thriving students, and intervention can give us actionable insights on students who show signs of potentially struggling in the course and beyond, early in a semester.

University of Notre Dame, is a medium-sized (a total of 8,530 undergraduate students by Fall 2016) private institution located in the Midwest U.S. The overall student body is 53% male and 47% female with 98% of students who began their studies in Fall 2015 returning in Fall 2016.

The 98% first-year retention rate makes it difficult to identify students who are not thriving at this university. However, the creation of a new First Year Experience (FYE) course in 2015 presented us with an opportunity to explore potential solutions. This course, now in its fourth year, is mandatory for all first-year students, of which over 1500 students are included in our analyses, and draws 125+ instructors, each of whom leads a standardized section of no more than 19 students. Consisting of two semester-long courses each worth one credit-hour and associated with a letter grade, the FYE helps students make a meaningful transition to collegiate life by integrating their academic, co-curricular, and residential experiences. As a mastery-based course, FYE is designed with the expectation that all students who put in the necessary effort should not only succeed but also be on a pathway to thrive. As a result, on an average, 90.00% of the students get an A as their final grade, with a standard deviation of 1.39% every semester.

In this paper, we demonstrate that our FYE course can give us insights into overall retention and the performance of students including all the classes they enroll in. Since the analyses include the majority of the first-year student body, subtle student behavior patterns that might often be overlooked in smaller classes can be more apparent. We also discuss our data pipeline process for capturing and analyzing all the data, our techniques to identify students who are not thriving in this introductory course, and our attempts at boosting them.

2 RELATED WORK

Learning Analytics is a field based on technology-enhanced learning [16] that focuses on the learning process [35]. In particular, it can greatly shape and impact learning in higher education [35]. While learning analytics can be deployed on many levels (e.g. department and institution), we focus on the course-level in this work, which is concerned with learning analytics deployed in classrooms [34]. Learning Analytics has been popularly used in institutions for student success and intervention [15], with a comprehensive list of the use-cases given in Dietz-Uhler *et al.* [15].

Recently, there is a growing emphasis on closing the learning analytics loop [12, 17, 25, 30] in which the results of predictive analytics and insights gleaned from them are used to improve the current or next iteration of a course in the form of interventions [12] and learning design [28]. In particular, Clow [12] recommends a five-step approach to this closed loop cycle: Capture, Report, Predict, Act, and Refine. We show in this paper how we use historical classroom data to improve our identification of non-thriving students in the next iteration of the course, thus closing the learning analytics loop. A recent example of this effort is by Choi *et al.* [11] who identify at-risk students using a simple metric and provide interventions to those students in one small course. In our work, we perform identification and intervention on the entire first-year body of students and repeat it for several semesters.

Every learning platform/institute has its own data collection and storage systems, and attempts to standardize these have not been widely successful [14]. In response, we propose a framework that can be tailored to build the underlying infrastructure. We aim to offer both reproducible steps that can be implemented in any classroom setting and to provide our work as evidence for the successful deployment of these cyclic steps.

First-year seminars, including our FYE course, have become increasingly popular. They are a high impact educational practice [23, 24] and their significance for retention, persistence, and engagement has been shown in the literature [19, 22, 27, 32]. Thus, we find it important to help students thrive in our FYE course. To generate the rich data on students' course activity in FYE (which is essential for actionable learning insights) and to promote active and student-centered learning, we took a flipped classroom approach [8, 9]. In our flipped FYE course, students participate actively in seminar-style discussions which build on their preparatory work at home.

The first step of identifying students that need to be boosted, the non-thriving students of a course, has been a popular area of research in the learning analytics community [7, 27, 29, 40]. Different data sources like demographic data, students' performance, and behavior are used to predict at-risk students. Some of these studies show improvements in student's grades after deploying these systems [7]. But, it is not clear if the improvement in learning outcomes is because of the intervention provided or if there were other factors involved because of a lack of evidence [17]. While these studies focus on at-risk students, we find the use of this term misleading in our case and potentially harmful as these students are not necessarily at risk of failing the class, but may struggle later or in other aspects of their campus life. In other words, our aim is not to help students survive, but to ensure that they thrive. We do this by including not only students at a risk of failing the course but also those in the bottom 2% of the course grades.

Once the non-thriving students are identified, various intervention strategies can be employed to improve the performance of these students. Some intervention strategies shift the effort to the students, with the system sending them an email [7], whereas other intervention mechanisms include intensive intervention within or outside the classroom [18]. Another commonly used approach is providing feedback to students using dashboards [13, 31, 33]. Our intervention strategy involves the campus support system in the form of academic advisors to directly intervene with the students, aided by diagnostic gradebook reports, to help identify and solve the problems that the students might be facing. The relationship between academic advising and student retention has been shown in [20, 36, 37]. In a later iteration of the course, we added a personalized action plan via email intervention. The progress of students can be monitored either at the end of the course [7] or throughout the course [18]. In our work, we tracked the progress of students at multiple points before the end of the semester, intervening regularly at mid-term and, in some semesters, earlier as well.

3 CONTEXT AND FRAMEWORK

3.1 Research Questions

Because the course is designed to provide a consistent environment for all students, it has easily accessible data and a controlled environment for research. In order to investigate the effectiveness of our approach, we outline research questions that help organize our analysis and evaluation of the strategy from multiple perspectives:

(1) RQ1 (Identification Criteria): How do we identify students who are not thriving and offer them support and encouragement to boost their success?
(2) RQ2 (Intervention Impact): What is the impact of our early and mid-semester analytics-based boost?
(3) RQ3 (FYE and Overall First Semester Performance): If the FYE is a common course for all students, could it serve as an indicator of overall first-year performance and retention?

3.2 Our Framework

We organize this paper according to our integrated Closed-loop Learning Analytics Scheme (iCLAS) shown in Figure 1. Section 4 (architecting for collection) describes the architecture of our system with "design", "build", and "capture" as its steps of actions. Section 5 (analyzing for action) describes our identification and intervention loop with "identify", "notify", and "boost" as its steps of actions. Section 6 (assessing for improvement) describes the effectiveness of the various components of our scheme with "evaluate" and "report" as its steps of actions.

The loop intersects in "evaluate" and "identify" due to our commitment to continuously improve our ability to identify and boost non-thriving students. We also close the loop between "report" and "design" by reporting our findings to the design team, so that

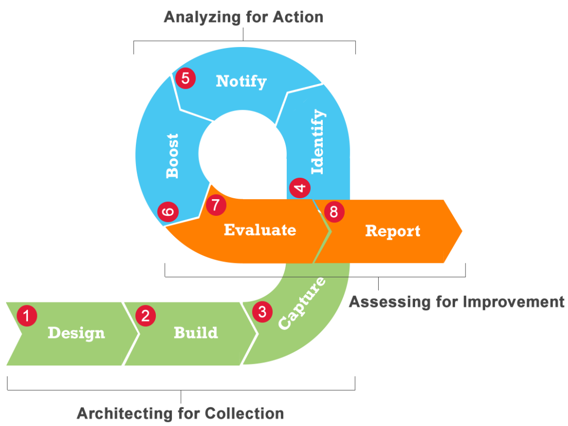

Figure 1: Integrated Closed-loop Learning Analytics Scheme

they may implement the required changes in the next iteration of the course. This iCLAS process should create a coherent and cohesive workflow that transcends courses and stakeholder perspective. Through this exposition, we hope that all stakeholders (program directors, instructors, advisors, students, data/learning scientists, and platform engineers) can recognize the design value of our integrative approach.

4 ARCHITECTING FOR COLLECTION

The first three steps in this foundational phase (Figure 1) optimize the opportunity for learning analytics by designing an engaging learning experience with standardized assessment and building a Next Generation Digital Learning Environment (NGDLE) that captures multidimensional data. In the first step, we *(1) design* an active, integrative student-centered learning experience for the course. With mastery learning in mind, we wanted the course to encourage critical, independent thinking for our students. In the next step, we *(2) build* a standard and integrated learning environment for the course. We wanted the environment to follow the NGDLE interoperability with integrative analytics, advising, and learning assessment principles in mind [5, 10]. Lastly, we ensure that our architecture has the capability to *(3) capture* student data from multiple sources in real-time into a centralized learning record warehouse as shown in Figure 2. With a centralized warehouse, we wanted to empower key stakeholders to make decisions based on actionable reports using real-time multidimensional data. These three steps ensure our ability to perform comprehensive analysis for action and conduct continuous assessment for improvement.

4.1 Design

Our primary design goal was to deliver an engaging and consistent learning experience to all FYE students and capture multidimensional data they generated in a centralized location in real-time to fuel actionable learning analytics. The goal was accomplished through an iterative and incremental development approach. The following section describes the approach and solutions in detail.

4.1.1 Overview of the Course Design.

Students meet in FYE sections for 50 minutes over 13 weeks of each semester. Before each session, students are provided with online materials to review and reflect on in a written weekly prompt assignment due before each class. In-person class meetings are discussion-based or active experiential learning on campus. After class, the weekly prompts are scored and students begin the process of preparing for the following week. At the mid- and end-point of each semester, students are given an in-class participation grade. Major assignments (integrations) occur twice a semester, at the mid- and end-points.

4.1.2 Assessment Design.

To ensure consistency, all 100+ course sections of FYE shared the same assessments and rubrics that consisted of weekly pre-class assignments, major integration assignments and participation grades both at the middle and end of the semester. Thus even though each section was graded by its own instructor, the students' grades were standardized and comparable across all the sections. Prior to each week's class, students were expected to complete a short reading/video viewing and write a 200-word response to a preparation prompt related to that material. The rubric had only 3 levels to create a simple and low-stakes scoring system for instructors to evaluate if students showed reasonable preparation (20 points), partial preparation (10 points), or no preparation (0 points). Prompts were designed primarily to hold students accountable for completion of the reading/viewing and prepare them to participate in discussions during the in-class meeting. Participation scores were assigned twice a semester. By providing participation scores at mid-term, students received feedback on their level of participation and could make a change, if necessary, for the second half of the semester. Multimedia ePortfolio assignments (integrations) were submitted three times a semester in the academic year of 2015-16, then reduced to twice a semester starting in Fall 2016. The same rubrics were used to assign scores in these categories to all students.

4.1.3 Standardized Grading & Gradebook.

Every graded item was scored from a universally-applied rubric by the instructor of the section. FYE program directors designed rubrics for weekly prompts, integrations, and participation as explained in Section 4.1.2. The use of common course grade-scales and identically constructed gradebook tools resulted in our ability to readily aggregate grade data from all sections and make direct comparisons and analyses across the entire first-year cohort of students.

4.2 Build

Our course design requires a standard and integrated learning environment. In order to implement such a learning environment, we followed the principles of NDGLE which focuses on bridging the gaps between current learning management tools and a digital learning environment that could meet the changing needs of higher education [10]. Sakai was chosen as the main hub for this learning environment, and we integrated all the tools required for course activities by following the interoperability and integration dimensions of NGDLEs [10]. Our integration process is iterative. We started from basic HTML iframe embeddings of videos and

Google Docs onto the course webpage and upgraded to advanced vendor-provided application programming interfaces (APIs). We eventually evolved to build Learning Tool Interoperability (LTI) solutions. LTI is a standard developed by the IMS Global Learning Consortium and aims to deliver a single framework for integrating any LMS product with any learning application [1]. The LTI integration not only allows students to perform all the required tasks in one central place but also enables the secure and trusted data flow between tools.

Another critical dimension of our learning environment is “analytics, advising, and learning assessment" [5, 10]. We intentionally built the environment to internally collect various sources of tool data like grades, click data, and ePortfolio assignment text. Our attempt to continuously improve the data collection process was iterative as well. We started with manually extracting clickstream and grades data in a batch periodically and upgraded to a Learning Record Warehouse (LRW) in real-time. The LRW was implemented based on the Apereo open-source learning record warehouse solution [6].

This upgrade removes the limitation of delayed on-time identification and assistance for non-thriving students presented in the batch process. In this upgraded system, every time a student performs a task in our learning environment, an xAPI or Caliper statement describing that experience is reported and stored in the LRW. For example, an experience is written as “student A performed action B with outcome C (in context D) at time E". xAPI is a new specification for learning technology that makes it possible to collect data about the wide range of experiences a person has (online and offline) [3]; Caliper offers the same ability with a richer set of specifications ("metric profile") [2]. Subsequently, the use of LRW solution resulted in the ability to record traces of student learning activity seamlessly in real-time. This eliminates the effort to manually extract data from each individual tool and makes real-time analytics possible. More importantly, this ambient data collection process does not impose any extra requirements on students.

4.3 Capture

Figure 2 describes the data collection process and pipeline. With the implementation of NGDLE and LRW, course activity data such as logging in and out, clicking on resources, attempting and submitting assignments were captured from Sakai in real-time in LRW. Time-on-task data such as the amount of time students actually spent on watching course videos were also collected from Panopto. This data revealed different aspects of students video watching behavior: how many times students viewed any given video, what segments of the video students selected to view, where did they stop viewing, what their average view rate was. Additionally, student performance indicators, such as their weekly prompts scores, ePortfolio integration scores, and class participation scores, were collected directly from the Sakai grades database to ensure data integrity and accuracy. This multidimensional data was merged in Tableau to develop insightful reports. Another reason we used Tableau is it makes it easier to share raw data or reports with different stakeholders. The data collection process and pipeline shown in Figure 2 is essential in our research and effort to boost every student’s potential to thrive.

5 ANALYZING FOR ACTION

From the first semester of FYE, we took steps to boost students towards positive educational outcomes. The three steps on this mindful action phase (Figure 1) are to identify students, notify them for action, and boost their success. We *(4) identify* students using a combination of learning design predictions (an educated guess of factors that show signs of students who are not thriving) and retroactive statistical analysis of students’ data that have been captured in the previous phase. Once we identified the students, we *(5) notify* them through two methods: bottom-up (inform and empower students via personalized action plan) and top-down (alert and empower advisors via one-on-one communication). In this process, we worked hard to prevent negative labeling of our students by not using words such as “at-risk" and “intervention." Instead, we adapted positive words such as “optimize", “boost", and “thrive." This leads us to the next step: to *(6) boost* the students’ success. We ensure that our boost from the student action plan and advising interactions is personalized based on an individual student’s circumstances. Ultimately, these three steps are designed to encourage student success in a more compassionate way.

5.1 Identify

In the first semester of Fall 2015, we identified two types of non-thriving behavior that resulted in early and mid-term boosts. The early boost was provided for students who had scores of 0 on their weekly prompts in weeks 2 and 3. The mid-term boost was for students who earned C- or lower at mid-term (Week 8 of 15 week semester) based on the institutional standard cutoff for Mid-Semester Deficiency Grade Reporting [4].

In the next three semesters (Spring 2016, Fall 2016, Spring 2017), we made two significant changes. First, we only identified the mid-term boost because we were struggling with the grade data reliability and data processing efficiency for analysis. Second, we adjusted our criterion for non-thriving to B- based on the grades distribution we observed in Fall 2015. Because most students get an A as their final grade (90 ± 1.39% on average), domain experts decided that a B- cutoff was more appropriate for identifying non-thriving students.

The data processing and reliability bottleneck was ameliorated by the implementation of the LRW in Spring 2017, which resulted in the automatic real-time data update [26]. With this improvement, an opportunity for an earlier identification of students who needed boost was presented to us. Therefore, we hypothesized (based on domain expertise) that students should be given an early boost when they showed no preparation (0 points) at least twice, either by not submitting weekly prompts or by submitting inadequate work, on assignments in between weeks 1 to 6 for Fall 2017.

5.2 Notify

The list of non-thriving students identified in the above section and grounds for their inclusion were shared with the FYE program director. For Fall 2015, the FYE program director notified the instructor of record in week 4 for the early boost. At mid-term, the FYE program director notified the first year advisors in week 8.

We changed our “notify" strategy over time to accommodate the requests and convenience of various stakeholders and incorporate the findings of the analysis in the previous semesters. In the next

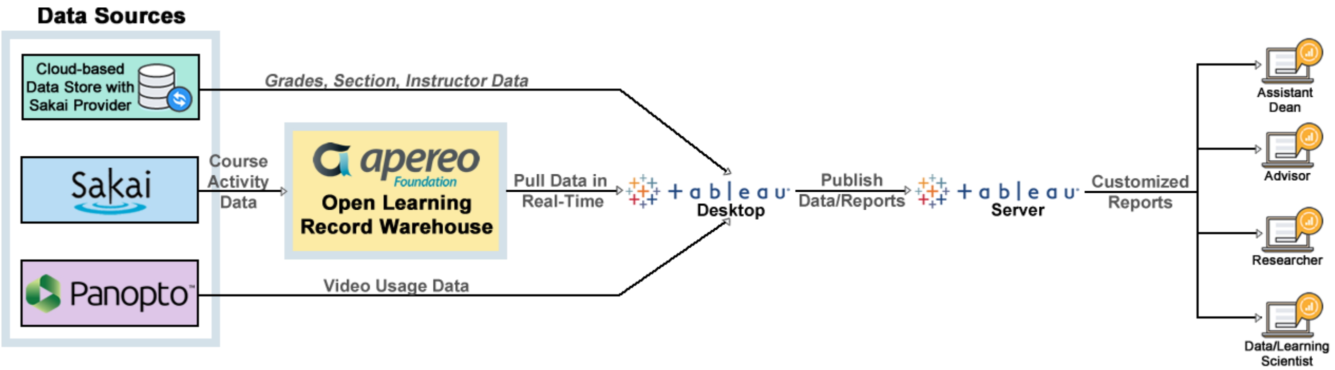

Figure 2: Platform Architecture

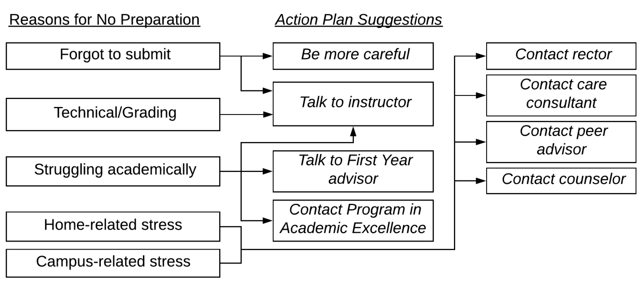

Figure 3: Bottom-up Method of Boosting Non-thriving Students

three semesters (Spring 2016, Fall 2016, Spring 2017), the FYE program director only notified advisors of non-thriving students at mid-term due to reasons including feedback from instructors and other stakeholders regarding the feasibility of this early intervention given instructors' workload. During the Fall 2017 semester, we added the early boost back and addressed the earlier concerns by empowering students to take direct action instead of relying on instructors and academic advisors to intervene. These non-thriving students received a message from the FYE program director to let them know that their behavior might be showing signs of struggling, to ensure that they had knowledge of resources, and to encourage them to choose a personalized action plan. We also notified the FYE program director and the instructor of record regarding these students.

5.3 Boost

In our first semester, instructors were encouraged to have conversations with students who were identified as part of the early boost. In week 8, first year advisors conversed with students who were identified at mid-term boost. The boost action for the next three semesters (Spring 2016, Fall 2016, Spring 2017) was solely data-driven discussion between students and their first-year advisors informed by diagnostic gradebook reports.

During Fall 2017, we re-packaged our early boost based on our analysis and the availability of all the necessary technology to implement it. We asked students to fill out a short qualtrics survey with tree-based logic to help them reflect on the reasons for their lack of preparation as reflected in their grades. Each reason led to a carefully selected list of recommended actions as shown in Figure 3, from which a student was asked to select their personalized action plan and avoid this situation in the future. The boost action for students at mid-term stayed the same as the other semesters.

6 ASSESSING FOR IMPROVEMENT

The last two steps in this continuous improvement phase (Figure 1) are to evaluate the impact of the course and report the findings to all stakeholders. Once we have acted on the students who were on the boosting list using the steps explained in Section 5, we *(7) evaluate* the intervention impact. Finally, through the architecture described in Section 4, we are able to *(8) report* insights into our data to multiple stakeholders using Tableau visualization. Administrators, instructors, advisors, and researchers benefited from the availability of reports on this data in order to analyze the trends of student engagement in the course.

6.1 Evaluate

Now that we have explained the course design and boost intervention strategy, we answer the research questions enumerated in Section 3.1.

Table 1: Odds Ratio

Semester	No preparation	Non-thriving	Thriving	Odds ratio
Fall 2015	≥ 2	4	44	10.4
	< 2	15	1718	
Spring 2016	≥ 2	17	38	27.0
	< 2	28	1687	
Fall 2016	≥ 2	10	27	27.5
	< 2	20	1483	
Spring 2017	≥ 2	8	29	27.2
	< 2	15	1480	

p-value < 0.005

6.1.1 RQ1: Identification Criteria.

Based on domain expertise, we used the midterm grade as a ground truth for non-thriving students. Intuitively, earlier boosts would help students solve the challenges they face earlier and thrive sooner. However, identification based on assignment scores has a bottleneck of scores being available only after instructors have graded the assignments and uploaded the scores. Therefore, we hypothesized that students who showed no preparation on their assignments (weekly prompts) at least twice within the early period of six weeks should be identified as non-thriving students and need to be boosted.

To verify this hypothesis, we retroactively analyzed the semesters of Fall 2015, Spring 2016, Fall 2016, and Spring 2017 to check if showing no preparation at least twice increased the risk of students having a B- (or C- in Fall 2015) or lower grade by mid-term (week 8 of 15). The criterion (no preparation) is not independent of the outcome variable (non-thriving mid-term grades) because mid-term grades are a summation of graded weekly prompts, integration, and participation scores up to week 8. In acknowledgement of this dependence and because both the criterion and the outcome are categorical variables, we use Fisher exact odds ratio test [21] to calculate the odds ratio between no preparation on at least two assignments and not-thriving. The null hypothesis is that the criterion does not affect the outcome. Table 1 shows the resulting contingency matrix. For each semester, the number of students under each category is listed, with the odds ratio. The null hypothesis can be rejected with a p-value < 0.005. Thus, we see that showing no preparation for at least two assignments affects the outcome of students being identified as non-thriving by mid-term.

Clickstream data could give us an even more fine-grained view of students' assignment submission patterns. Therefore, we checked for correlation between the clickstream data of students who had a non-thriving grade by mid-term, but the results were inconclusive. We also considered using integration and participation scores, but they were populated very close to the mid-term point and were not early enough indicators for non-thriving behavior. Hence, we decided to use showing no preparation (indicative of no submission or a score of zero) on at least two assignments as an early indicator for non-thriving students.

Table 2: Confusion Matrix for Fall 2017's Early Intervention

	Grade until Week 6	
No preparation	**$\leq$ B−**	**$\geq$ B**
≥ 2	14	17
< 2	14	1676

In the Fall of 2017, we used this criterion as an early indicator for non-thriving students since all the required technology to implement this was finally in place. To assess its effectiveness , we show a confusion matrix with students having a B- or lower by week 6 as the ground truth. We should not use their mid-term grades as the ground truth because the intervention may interfere with their grades and change the mid-term grade. Since the early intervention occurs after week 6, the grades calculated until week 6 would not be affected and can be used as ground truth. Table 2 shows the associated confusion matrix. Because a very small proportion of students were non-thriving, even if all the non-thriving students were wrongly identified, we would have a high accuracy. Thus, a 98.1% accuracy is misleading as a performance metric. Instead, we used Cohen's Kappa, which is commonly used for measuring inter-rater agreement. In our case, one rater, the oracle, can look into the future after the assignments have been graded and knows the ground truth of which students have a B- or lower at week 6. The second rater sees only the past criterion of having at least two grades indicating no-preparation. We computed Cohen's Kappa to check how much the past criterion agrees with the oracle, as a measure of the effectiveness of the second rater. The calculated Cohen's Kappa score is 0.4654 which is generally accepted to show moderate agreement [38] between our criterion and the oracle. Thus, showing no preparation on at least two assignments is a moderately reasonable criterion for identifying non-thriving students.

In order to examine the relationship between non-thriving students and their assignments scores, we studied the correlation between them for each semester individually. We also combined the Fall and Spring semesters to see dominating patterns. Table 3 shows the point-wise biserial correlation coefficients for the Fall and Spring semesters. We see that all weekly scores are significantly correlated in all the semesters except Fall 2015 and Spring 2018, with a p-value < 0.05. Each semester has a different ranking of the weekly assignments depending on the correlation with non-thriving students. This ranking is not consistent over the semesters. Moreover, the differences between the correlation coefficients is not drastic. This seems to imply that all the weekly prompt grades are approximately equally important for identifying the non-thriving students.

6.1.2 RQ2: Intervention Impact.

To evaluate the effectiveness of our interventions, we compared the change in performance of students between those who were boosted (intervention group), and those who were not (control group). This is not a truly randomized control/intervention division because the intervention group consisted entirely of all the non-thriving students, and each student in the intervention group has a lower grade than the students in the control group. To measure if the change in the outcome variable is statistically significant, we used a pre-post test with paired data. Specifically, we utilized a one-tailed non-parametric pre-post test, the Mann-Whitney U test, because the the grades of students is not normally distributed, with most of the students getting full scores. Table 4 shows the mean and standard deviation of different groups of students. The change in grade was computed by subtracting the mid-term grade from the final grade for each student. The Mann-Whitney U test showed that the students in the intervention group had a significantly higher change in grade compared to the control group with a p-value < 0.0001. The reported means and standard deviations of the two groups showed that the difference between them is huge. Generally, our results are consistent across the semesters. Moreover, the majority of non-thriving students (73.6%-87.6%) improved their grades between mid-term and final each semester.

While these results seem encouraging, the non-intervention group has a large fraction of students with A's in them, and these students have a very small scope of improvement compared to the students in the intervention group. To reduce the mean difference in

Table 3: Weekly Scores Correlation with Non-thriving Students for Fall and Spring Semester

Fall 2015		Fall 2016		Fall 2017		Fall Combined		Spring 2016		Spring 2017		Spring 2018		Spring Combined	
Scores	CC	Scores	CC	Scores	CC	Scores	Corr. Coeff.	Scores	CC	Scores	CC	Scores	CC	Scores	CC
Week 5	-0.165	Week 7	-0.252	Week 4	-0.344	Week 4	-0.140	Week 5	-0.317	Week 6	-0.246	Week 2	-0.238	Week 6	-0.212
Week 3	-0.123	Week 1	-0.231	Week 5	-0.259	Week 5	-0.128	Week 1	-0.312	Week 4	-0.193	Week 4	-0.206	Week 1	-0.196
Week 2	-0.0832	Week 6	-0.205	Week 7	-0.253	Week 6	-0.0812	Week 6	-0.214	Week 3	-0.191	Week 6	-0.170	Week 4	-0.192
Week 6	-0.0670	Week 5	-0.165	Week 3	-0.222	Week 3	-0.0782	Week 4	-0.184	Week 2	-0.186	Week 1	-0.169	Week 2	-0.190
Week 4	-0.0496	Week 4	-0.0929	Week 6	-0.198	Week 2	-0.0710	Week 2	-0.174	Week 5	-0.170	Week 3	-0.146	Week 5	-0.178
Week 1	-0.0146†	Week 2	-0.0873	Week 1	-0.166	Week 7	-0.0699	Week 3	-0.142	Week 7	-0.169	Week 7	-0.125	Week 3	-0.158
Week 7	-0.00161 †	Week 3	-0.0686	Week 2	-0.146	Week 1	-0.0541	Week 7	-0.0638	Week 1	-0.0510	Week 5	-0.0421 †	Week 7	-0.110

p-value < 0.05 except where † : $p - value > 0.05$ (not significant)

Table 4: Improvement in FYE grades compared between students who do and do not receive intervention

Semester	Grade change intervention Mean ± Std. Dev.	Grade change no intervention Mean ± Std. Dev.	p-value	Achievement ratio intervention Mean ± Std. Dev.	Achievement ratio no intervention Mean ± Std. Dev.	p-value
Fall 2015	2.519 ± 1.056	0.0368 ± 0.275	*	1.170 ± 0.351	0.994 ± 0.0620	*
Spring 2016	0.873 ± 0.959	0.0447 ± 0.259	*	0.939 ± 0.339	0.999 ± 0.0601	–
Fall 2016	1.298 ± 1.310	0.0403 ± 0.261	*	0.997 ± 0.380	0.997 ± 0.0500	–
Spring 2017	1.318 ± 1.195	0.0183 ± 0.187	*	1.043 ± 0.325	0.997 ± 0.0414	†
Fall 2017	0.682 ± 1.137	0.0504 ± 0.201	*	0.985 ± 0.299	1.003 ± 0.0315	–
Spring 2018	1.123 ± 1.428	0.0244 ± 0.193	*	1.039 ± 0.383	0.998 ± 0.0399	–

The p-value is calculated using one-tailed Mann-Whitney U test. Legend: p-value < 0.0001 :*, p-value < 0.01 :†, p-value > 0.05 (not-significant):–

mid-term grades of these groups, we considered a smaller subset of students in the non-intervention group. In Fall 2015, only students with a C- or below were boosted, as opposed to B- and below for all the future semesters. This gave us the opportunity to use the students within the range of B- and C- as a group of students against which we can compare the performance of the intervened students. Once again, this is not a randomized control group, because the students in this group start out with a higher grade than the students in the intervention group. We will refer to this group as the B- to C- control group henceforth. However, we can compare the change in the grade of the students between the mid-term and the end of the semester. This result can give us some indication of the effectiveness of our intervention. The number of students with a C- or less that received an intervention is similar to the number of students in our B- to C- control group. An unpaired one-tailed Mann-Whitney U test between the changes in mid-term to final grades of the two groups showed that the intervention group had a statistically significantly greater change in grade, with a p-value < 0.0001. In fact, the mean change of grade for the B- to C- control group was 0.795, with a standard deviation of 0.607, whereas the mean change of grade for the intervention group was 2.61, with a standard deviation of 1.10.

While students in the intervention group in Table 4 improve their grades significantly more than the original control group, the students in the control group do not have as much scope for improvement as the students in the intervention group. To mitigate this, we calculate the achievement ratio to measure the potential that a student reaches compared to the maximum possible grade they can achieve, instead of measuring the change in grade. This is calculated by:

$$\text{achievement ratio} = \frac{\text{final grade}}{\text{max. possible grade}}, \text{ where}$$
$$\text{max. possible grade} = \text{mid term component} \cdot \text{mid term grade} + (1 - \text{mid term component}) \cdot 4$$

The maximum possible grade is weighted by the mid-term component of the grade, which denotes the ratio of the contribution of the mid-term grade to the final grade. Table 4 shows the mean and standard deviation of the achievement ratio of the students, both in the intervention group and control group, in each semester. The achievement ratio of the intervention group is not statistically significantly greater than the control group, except in Fall 2015 and Spring 2017. We see that the effect size is small from the reported means and standard deviations of the two groups. Since we do not have a randomized control group, we cannot know definitely whether the lack of differences we see is because the intervention does not have a long-term effect, or that there were other factors related to the grades of students (e.g. internal motivation). We also note that in some cases, the average achievement ratio is greater than 1. Many students do, in fact, get a final grade that is greater than their maximum possible grade. This anomaly comes from grade reporting errors and grades being added or modified later in the semester by instructors. While we can design systems to keep track of grades entered in real-time, ultimately, on-time correct grade entry is still in the hands of the instructors. We will consider ways to reduce this problem in the future.

We can also track the set of non-thriving students from Fall to Spring semester. The academic years of 2015-16 and 2016-17 had two mid-term interventions performed in each year, with one in each of the Fall and Spring semesters. To evaluate whether students

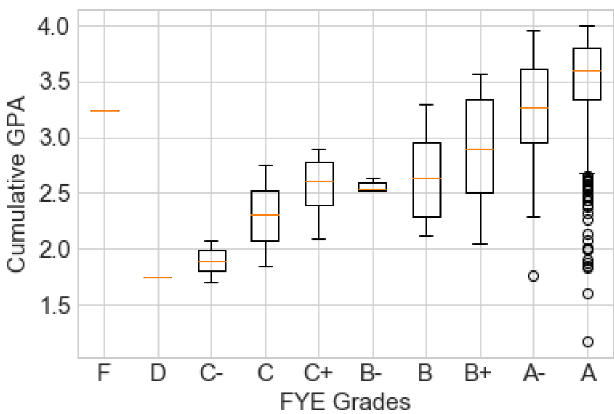

Figure 4: Spring 2017 - Final FYE Grades plotted against Cumulative GPA

who were boosted stay boosted, we looked at how many students identified in the Fall semester were again identified as non-thriving in the Spring semester. Less than 15% of the students identified as non-thriving students were identified again in the corresponding Spring semester for all the years. Thus we see very little overlap between the Fall and Spring non-thriving students.

6.1.3 RQ3: FYE and Overall First Semester Performance. In this section, we explore the impact of students' performance in FYE beyond the scope of FYE. Specifically, we find the relationship between students' performance in FYE and their overall performance and retention in the first-year.

Table 5: Correlation of FYE Final Grades with Cum. GPA and Cumulative GPA Differences between Non-thriving and Thriving Students for each semester

Semester	Pearson r	Cum. GPA non-thriving students Mean ± Std. Dev.	Cum. GPA thriving students Mean ± Std. Dev.
Fall 2015	0.386	2.76 ± 0.88	3.44 ± 0.44
Spring 2016	0.408	2.83 ± 0.76	3.45 ± 0.41
Fall 2016	0.321	2.78 ± 0.60	3.49 ± 0.40
Spring 2017	0.350	2.82 ± 0.51	3.50 ± 0.39
Fall 2017	0.325	2.97 ± 0.91	3.54 ± 0.43
Spring 2018	0.250	2.85 ± 0.68	3.54 ± 0.37
p-value < 0.0001		p-value < 0.0001	

Table 5 shows significant positive correlation between the FYE grade and cumulative GPA of students with p-values < 0.0001. For illustration purposes, Figure 4 shows an example from Spring 2017, where for each FYE final grade (x-axis), the boxplots of the cumulative GPA corresponding to those students is plotted (y-axis). Thus, even though the FYE is only a 1 credit course of the minimum of 12 credits a full-time student takes in a semester, we find a consistently positive correlation between the FYE grade and cumulative GPA for all the semesters. This indicates that the performance in the FYE course can provide insights into the students' overall performance.

Table 6: Correlation of weekly homework with retention

Academic Year 2015-16 Feature	CC	Academic Year 2017-18 Feature	CC
Week 5	-0.109	Week 1	-0.0890
Week 6	-0.0961	Week 5	-0.0814
Week 3	-0.0729	Week 3	-0.0779
Week 7	-0.0676	Week 4	-0.0733
		Week 6	-0.0727
non-thriving students	0.137	non-thriving students	0.0648
p-value < 0.01			

To evaluate whether our identification criteria is effective beyond FYE, we compared the cumulative GPA of non-thriving students with those who are thriving for each semester. Table 5 shows the mean and standard deviation for these groups per semester. The thriving and non-thriving students have statistically significant differences in cumulative GPA every semester, with a p-value < 0.0001 using the non-parametric Mann-Whitney U test.

The issue of retention can be investigated by observing the behavior of students who are no longer with the university. The students who withdraw, are dismissed, apply for a leave of absence, or are suspended, comprise this set. By the time we identify these students within the semester, it is often too late. To identify these students earlier, we examined the correlation between the grades of students in weeks 1-7 of the Fall semester with their enrollment status in the Spring semester as the y-variable. We restricted our analysis to assignments before the intervention, because the intervention may affect the student's retention. Since the y-variable has only two values, enrolled and dismissed, we once again used point-wise biserial correlation. We do not observe significant correlations for the year of 2016-17, but some weeks are correlated with the retention of students in 2015-16 and 2017-18, shown in Table 6. There is also a slight significant correlation with non-thriving students.

6.2 Report

The regularized and multidimensional data enabled us to perform high-level analysis and develop insightful reports to help FYE senior administrators make data-informed strategic decisions. We used Tableau, a business intelligence and analytics tool, to merge the activity and performance data from multiple sources, perform aggregate analyses, and create intuitive and insightful visualization reports. The final reports were shared with different stakeholders, such as assistant deans, program administrators, advisors, and researchers, through our Tableau server. The reports were designed to offer insights on various aspects of the FYE program. For example, to facilitate the continuous improvement of course design and provision of learning materials, we built reports to show which learning materials were most engaging and what was the optimal timing for selected materials. Based on the reports, the course design team removed the reading materials that were less engaging and adjusted the video materials to the optimal length. These strategies would help improve student engagement through better course design. We also built reports to highlight the frequency of non-submissions on assignment grouped by student, assignment, and section. These reports helped program directors monitor the progress of the course and identify opportunities to stimulate student performance. Additionally, we built program-wide grade distribution reports to empower instructors to measure and adjust their own grading practices, answer student questions on whether they are graded fairly, and help advisors to develop a holistic view of their advisees' scores. All the reports were updated and shared on a weekly basis so that FYE administrators would have the most timely information on how to continuously improve the program's effectiveness, and enhance student success and satisfaction rate.

7 DISCUSSION

A limitation of our evaluation is the lack of a randomized control group. To study the long-term effects of the intervention on the performance of students in FYE, we tracked the performance of the Fall intervention and control group through the Spring semester by measuring the change between their Spring mid-term and final FYE grades. However, we did not find the intervention group to have a higher change in grades compared to the control group. This may suggest that even if our intervention has short-term effects of improving students' grades, it might not translate to a long-term performance improvement. To establish the intervention as the cause for students' grades improving, we need a randomized control group of students who do not receive the intervention. However, at the time of designing the course and intervention strategy, it was deemed unethical and unfair to provide some students with extra resources and assistance while depriving others to form a control group. This is the conundrum of impactful intervention research in real-world instead of a controlled lab setting. Finding an ethical way to provide intervention for all students that appear to need a boost while providing a control group as a way to conclusively establish the intervention as the cause of students' improvement in grades will be part of our future pathway.

We initiated an early-boost in Fall 2015 limited to students who missed assignments in both weeks 2 and 3. This was a design decision as opposed to a data-driven decision because students are allowed to switch sections in week 1 of the course. When students switch sections, the grades from the previous section are not transferred automatically. Hence to ensure a more stable population, weeks 2 and 3 were chosen to be the indicator for students who needed an early-boost. With these limitations, we saw an opportunity for improvement. We will also extend the definition and identification of non-thriving students to include indicators besides grades, e.g., clicks, cumulative GPA, the trends of students' grades as opposed to absolute grades, and other non-academic factors.

The mutually iterative relationship we developed between the course design and data collection/analysis helped us continuously improve the student learning experience and enhance our effort to help every student succeed. Learning scientists took the lead on hypothesis, feature predictions, and interventions. Blended into the process, the data scientists evaluated and refined identification and boosting methods. The goal of this process was to continually improve the analytics-based strategy by mining multiple years of historical data.

As mentioned in Section 4, the upgrade to LRW resulted in the ability to do real-time analytics and the capacity to do a more refined early boost for non-thriving students. In Fall 2017, we hypothesized that students should be given an early boost when they showed no preparation at least twice in weeks 1-6. This design-based decision was backed by multiple analyses as shown in Section 6.1.1. There were more false positives and false negatives compared to true positives (students who are on the early-boost list) in the odds ratio analysis (Table 2). This may be because of the needle-in-a-haystack nature of finding non-thriving students. We hope to iterate more on the identification of early-boost list students and reduce the number of false positives and false negatives in the future.

We proposed that showing no preparation at least twice on weekly prompts as a more consistent indicator of non-thriving students. This means that we might consider notifying and boosting students automatically as soon as they miss two of their assignments instead of waiting for arbitrary 1/3 and 1/2 semester cutoffs. However, we would also need to be cautious and sensitive. We want to enhance students' ability to succeed instead of labeling them as at-risk. We want to nudge them into successful student behaviors instead of criticizing their inability to complete their assignments. We see First Year Experience as the appropriate course to start this endeavor. We have shown that there is a correlation between students' first semester cumulative GPA and their performance in the course in Section 6.1.3. It is also a standard learning experience with less variability than other courses taught on campus. Along the way, we have developed the integrated closed-loop learning analytics scheme that consists of the backend NGDLE infrastructure, data pipelines, strategies to notify and boost students, and the front-end stakeholders interface. We hope that the NDGLE infrastructure established to capture, visualize, and analyze the data can be adapted to other large credit-granting courses.

8 CONCLUSION

We examine our research endeavor using three probing research questions that deal with our goals to effectively identify and boost students who were not thriving in a timely manner.

Our first research question deals with ways to accurately identify a small proportion (2% of the total 2,000 first year class) of non-thriving students without harm and as early as possible. We accomplished this through capturing the data they generated in real-time and performing analysis from multiple perspectives. Using various statistical methods, we can see moderate correlation between non-thriving students and no preparation on at least two assignments six weeks into the semester. However, improvements can be made in the future to reduce error in classification.

Our second research question quantifies the impact of our early and mid-semester boost. Using Mann-Whitney U analysis, we see a significantly higher change in grade but not achievement ratio for the boosted students. We also see little overlap between non-thriving students identified in the fall and those identified in the spring. However, since we do not have a randomized control group, it is challenging to establish our intervention as the cause of non-thriving students' improvement in performance. We hope to find an ethical way to do so in the future.

The third research question investigates the impact of the FYE course on students' overall First Year grade performance and retention. We see a significant positive correlation between the FYE grade and their cumulative GPA. This is consistent with the literature that shows the impact of First Year Experience courses with respect to retention, persistence, and engagement. We will continue our investigation to understand FYE's relationship to other introductory courses commonly taken in the first year and retention.

This integrated closed-loop learning analytics scheme (iCLAS) goes beyond retroactive analytics. The scheme collects digital learning data using the Next Generation Digital Learning Environment. It takes action in real-time to boost students based on both design and data-driven insights. It evaluates its impact for continuous improvement and provides reports for multiple stakeholders in real-time or between iterations. It utilizes a First Year Experience course with standardized assessment and rubrics that provide fast and frequent low-stakes weekly assignments. This enables us to provide effective ways to obtain a real-time pulse of the students and encourages them to thrive in their first year of higher education.

9 ACKNOWLEDGEMENTS

We thank everyone in our team for the last three years, especially, Kevin Abbott, Kevin Barry, Chris Clark, Hugh Page, Maureen Dawson, Patrick Miller, Sharif Nijim, and Erin Hoffmann Harding.

REFERENCES

[1] [n. d.]. Basic Overview of How LTI Works. ([n. d.]). https://www.imsglobal.org/basic-overview-how-lti-works
[2] [n. d.]. Caliper Analytics. ([n. d.]). http://www.imsglobal.org/activity/caliper
[3] [n. d.]. What is the Experience API? ([n. d.]). https://xapi.com/overview/
[4] 2014. (2014). https://web.archive.org/web/20141021223329/http://facultyhandbook.nd.edu:80/assets/137497/undergraduate_academic_code_final_for_7.1.2014.docx
[5] G. Alex Ambrose, Kevin Abbott, and Alison Lanski. 2017. Under the Hood of a Next Generation Digital Learning Environment in Progress. (Jul 2017). https://er.educause.edu/articles/2017/7/under-the-hood-of-a-next-generation-digital-learning-environment-in-progress
[6] Apereo-Learning-Analytics-Initiative. 2018. Apereo Learning Analytics Initiative - OpenLRW. (Jul 2018). https://github.com/Apereo-Learning-Analytics-Initiative/OpenLRW
[7] Kimberly E. Arnold, Matthew D. Pistilli, and Kimberly E. Arnold. 2012. Course Signals at Purdue: Using Learning Analytics to Increase Student Success. *2nd International Conference on Learning Analytics and Knowledge* May (2012), 2–5. https://doi.org/10.1145/2330601.2330666 arXiv:0709.1706v2
[8] Jonathan Bergmann and Aaron Sams. 2012. Flip your classroom: Reach every student in every class every day. (2012).
[9] Jacob Lowell Bishop, Matthew A Verleger, et al. 2013. The flipped classroom: A survey of the research. In *ASEE national conference proceedings, Atlanta, GA*, Vol. 30. 1–18.
[10] Malcolm Brown, Joanne Dehoney, and Nancy Millichap. 2015. The Next Generation Digital Learning Environment: A Report on Research. (Apr 2015). https://library.educause.edu/resources/2015/4/the-next-generation-digital-learning-environment-a-report-on-research
[11] Samuel PM Choi, Sze Sing Lam, Kam Cheong Li, and Billy TM Wong. 2018. Learning Analytics at Low Cost: At-risk Student Prediction with Clicker Data and Systematic Proactive Interventions. *Journal of Educational Technology & Society* 21, 2 (2018), 273–290.
[12] Doug Clow. 2012. The learning analytics cycle. In *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge - LAK '12*. 134. https://doi.org/10.1145/2330601.2330636
[13] Linda Corrin and Paula De Barba. 2014. Exploring students ' interpretation of feedback delivered through learning analytics dashboards. *Proceedings of ascilite Dunedin 2014* November (2014), 629–633.
[14] Angel Del Blanco, Angel Serrano, Manuel Freire, Ivan Martinez-Ortiz, and Baltasar Fernandez-Manjon. 2013. E-Learning standards and learning analytics. Can data collection be improved by using standard data models?. In *IEEE Global Engineering Education Conference, EDUCON*. 1255–1261. https://doi.org/10.1109/EduCon.2013.6530268
[15] Beth Dietz-Uhler and Janet E Hurn. 2013. Using learning analytics to predict (and improve) student success: A faculty perspective. *Journal of Interactive Online Learning* 12, 1 (2013), 17–26.
[16] Rebecca Ferguson. 2012. Learning analytics: drivers, developments and challenges. *International Journal of Technology Enhanced Learning* 4, 5-6 (2012), 304–317.
[17] Rebecca Ferguson and Doug Clow. 2017. Where is the evidence?. In *Proceedings of the Seventh International Learning Analytics & Knowledge Conference on - LAK '17*. ACM, 56–65. https://doi.org/10.1145/3027385.3027396
[18] Douglas Fuchs and Lynn S Fuchs. 2006. Introduction to response to intervention: What, why, and how valid is it? *Reading research quarterly* 41, 1 (2006), 93–99.
[19] Kathleen Goodman. 2006. First-year seminars increase persistence and retention. *First-Year Programs* (2006).
[20] Deryl K Hatch and Crystal E Garcia. 2017. Academic advising and the persistence intentions of community college students in their first weeks in college. *The Review of Higher Education* 40, 3 (2017), 353–390.
[21] Eric Jones, Travis Oliphant, Pearu Peterson, et al. 2001. SciPy: Open source scientific tools for Python. (2001). http://www.scipy.org/
[22] Cindy A Kilgo, Jessica K Ezell Sheets, and Ernest T Pascarella. 2015. The link between high-impact practices and student learning: Some longitudinal evidence. *Higher Education* 69, 4 (2015), 509–525.
[23] George D. Kuh. 2008. *High-impact educational practices: what they are, who has access to them, and why they matter.* Association of American Colleges and Universities. https://www.aacu.org/publications-research/publications/high-impact-educational-practices-what-they-are-who-has-access-0
[24] George D. Kuh. 2009. What Student Affairs Professionals Need to Know About Student Engagement. *Journal of College Student Development* 50, 6 (2009), 683–706. https://doi.org/10.1353/csd.0.0099
[25] Karen D Mattingly, Margaret C Rice, and Zane L Berge. 2012. Learning analytics as a tool for closing the assessment loop in higher education. *Knowledge Management & E-Learning: An International Journal (KM&EL)* 4, 3 (2012), 236–247.
[26] Patrick Miller and Xiaojing Duan. 2018. NGDLE Learning Analytics: Gaining a 360-Degree View of Learning. (Jan 2018). https://er.educause.edu/blogs/2018/1/ngdle-learning-analytics-gaining-a-360-degree-view-of-learning
[27] Sunny Moon. 2013. High-Impact Educational Practices as Promoting Student Retention and Success. (2013), 203–221.
[28] Yishay Mor, Rebecca Ferguson, and Barbara Wasson. 2015. Editorial: Learning design, teacher inquiry into student learning and learning analytics: A call for action. *British Journal of Educational Technology* 46, 2 (2015), 221–229. https://doi.org/10.1111/bjet.12273
[29] Paul A Murtaugh, Leslie D Burns, and Jill Schuster. 1999. Predicting the Retention of University Students. *Research in Higher Education* 40, 3 (1999), 355–371. https://doi.org/10.1023/A:1018755201899
[30] Abelardo Pardo, Roberto Martínez-Maldonado, Simon Buckingham Shum, Jurgen Schulte, Simon McIntyre, Dragan Gašević, Jing Gao, and George Siemens. 2017. Connecting data with student support actions in a course: A Hands-on Tutorial. In *Proceedings of the Seventh International Learning Analytics & Knowledge Conference on - LAK '17*. 522–523. https://doi.org/10.1145/3027385.3029441
[31] Yeonjeong Park and Il-Hyun Jo. 2015. Development of the Learning Analytics Dashboard to Support Students ' Learning Performance Learning Analytics Dashboards (LADs). *Journal of Universal Computer Science* 21, 1 (2015), 110–133. https://doi.org/10.3217/jucs-021-01-0110 arXiv:110-133
[32] Stephen R Porter and Randy L Swing. 2006. Understanding how first-year seminars affect persistence. *Research in Higher Education* 47, 1 (2006), 89–109.
[33] Jose Luis Santos, Katrien Verbert, Sten Govaerts, and Erik Duval. 2013. Addressing learner issues with StepUp! *Proceedings of the Third International Conference on Learning Analytics and Knowledge - LAK '13* (2013), 14. https://doi.org/10.1145/2460296.2460301
[34] George Siemens. 2013. Learning analytics: The emergence of a discipline. *American Behavioral Scientist* 57, 10 (2013), 1380–1400.
[35] George Siemens and Phil Long. 2011. Penetrating the fog: Analytics in learning and education. *EDUCAUSE review* 46, 5 (2011), 30.
[36] Hadyn K Swecker, Matthew Fifolt, and Linda Searby. 2013. Academic advising and first-generation college students: A quantitative study on student retention. *NACADA Journal* 33, 1 (2013), 46–53.
[37] Jörg Vianden and Patrick J Barlow. 2015. Strengthen the bond: Relationships between academic advising quality and undergraduate student loyalty. *The Journal of the National Academic Advising Association* 35, 2 (2015), 15–27.
[38] Anthony J Viera, Joanne M Garrett, et al. 2005. Understanding interobserver agreement: the kappa statistic. *Fam Med* 37, 5 (2005), 360–363.
[39] Alyssa Friend Wise. 2014. Designing pedagogical interventions to support student use of learning analytics. In *Proceedings of the fourth international conference on learning analytics and knowledge*. ACM, 203–211.
[40] Annika Wolff, Zdenek Zdrahal, Andriy Nikolov, and Michal Pantucek. 2013. Improving retention: Predicting at-risk students by analysing clicking behaviour in a virtual learning environment. *LAK '13 Proceedings of the Third International Conference on Learning Analytics and Knowledge* (2013). https://doi.org/10.1145/2460296.2460324