2019

Smart Learning Analytics and Frequent Formative Assessments to Improve Student Retention

Hassan, Muhammad Mustafa

IEEE

Artikkelit ja abstraktit tieteellisissä konferenssijulkaisuissa
© IEEE
All rights reserved
http://dx.doi.org/10.1109/SMARTNETS.2018.8707435

https://erepo.uef.fi/handle/123456789/7681
Downloaded from University of Eastern Finland's eRepository
Smart Learning Analytics and Frequent Formative Assessments to Improve Student Retention

Muhammad Mustafa Hassan1,2, Adnan N. Qureshi1, Andrés Moreno2, Markku Tukiainen3
1Faculty of Information Technology, University of Central Punjab, Lahore, Pakistan
2School of Computing, University of Eastern Finland, Joensuu, Finland
3{Mustafa.Hassan, Dr.Qureshi} @ucp.edu.pk, {amoreno, Markku.Tukiainen} @cs.uef.fi

Abstract—In today’s world of competitive educational institutions, it is imperative that the final product, the students, are of the optimal quality as required by the professional industry. Conventionally, the progress and quality of the students were only assessed in end-of-term results, or at a few standpoints, which neglected the possibility of improving the weak areas. This shortcoming of the conventional educational system resulted in a high rejection/dropout of the potentially capable students. In this work, the authors propose adaptation of a novel content delivery, formative assessment, smart analytics and instant feedback mechanism pipelined into the educational process. The proposed model can potentially circumvent the pitfalls and significantly reduce the errors of assessment, grading, and the delivery of feedback. The proposed approach concurrently assures the quality of students at each formative step within a semester’s time, thus improving the quality of intake of the subsequent standpoint. The approach has been evaluated on one subject, Functional English, within a four-year Computer Science Baccalaureate Program. The results of the outcomes can be plausibly extended and applied onto other educational contexts.

Keywords—Smart Analytics, In-situ Analytics, Smart Learning and Assessment Environments, Closed Feedback, Formative Assessment, Educational Data Visualization.

I. INTRODUCTION

The withdrawal (dropout) of students have been a critical issue since early days of modern education systems. To name a few, the inability to assess the current learning situation, the inability of reconciling, financial crisis, moving away to another place, are some notable problems [1]. Nonetheless, a few of the aforementioned are out of the control of the academic system, but a few others are the areas needing improvement [1]. One such area is formative assessment and closed feedback loops delivering instantaneous information to all stakeholders.

Conventionally, the educational process has been designed to assess the quality of learning at a few designated standpoints throughout the semester. A number of courses assess the students twice in a semester in form of summative mid-term and final-term exams, with a few—two to five mostly—formative, relatively smaller, low-stake assessment instruments, administered in between [2]. Administering assessments beyond this number is mostly considered bothersome by teachers, as well as by students [2]. It is especially burdensome for teachers working with large enrolments and/or offering multiple courses in parallel—a case highly likely in educational systems of developing nations [3], as well as a growing issue in industrialized world [4]. Thus, the number of assessments—in any form, ranging from quizzes to exams, and assignments to presentations—is kept to humanly limit.

The number of assessments and the timely delivery of feedback is directly related to the quality of learning [4]. In a model with summative assessment only, or with few formative assessments only, a carryover effect is observed between learning modules. When large periods of learning are gone un-tested, an error in the previous learning module is undiscovered, and carried over to the next module to distort further learning. Consequently, when the error is discovered, it is often too late to go back and correct the stem, and all faulty branches [4]. The students are already at the verge of the final assessment. This results in a significant number of students dropping out or failing in the course, and in subsequent courses as well.

Contrary to the aforementioned summative educational model, the humans have received a major paradigm shift in industrial process in the recent decennia. The introduction of the assembly pipelines have allowed mass production and timely delivery of high quality products with minimized chances of delivering faulty artefacts. To allow mass production with stringent quality control, the assembly lines break down the entire process into its simplest atomic tasks with localized quality control loop. The pipeline at each grained step ensures the quality of the passing artifact, and does not allow a transition if the quality is not met. The artifact re-enters the same stage if quality problems are found, and keeps in local loop until the quality requirement of the transition are not met. Since the faults are tracked and tackled at early stages, or as soon as they are introduced, there is a high chance of correction. The problematic artifact does not travel farther in pipeline, and hence the error effect is not accumulated. This results in a significantly reduced number of faulty artefacts skipping out to the end of the pipeline, minimizing the chances of producing failures/dropouts.

The expected benefits of the proposed system are multi-fold. First, the localized quality assurances ensures timely discovery of errors. Second, instant feedback improves learning effectiveness: students—as well as teachers and parents—are informed about lacking areas, and hence, improvement possibilities in appropriate timeframe are created. Third, in-situ processing keeps all stakeholders in a continual informed state via smart analytics. Fourth, automated grading ensures fewer human-induced errors when marking response, as well as reduces teachers’ load, allowing more time spent on remedial measures for erroneous knowledge areas. Fifth, frequent testing increases knowledge retention for further summative assessments [2]. Summarizing, the local correction of the erroneous artifacts,
and the avoidance of carryover to further stages is expected to result in significantly less failures/dropouts.

Implementing and running such coarse delivery, assessment, grading, and feedback system manually is laborious, if not impossible. Fortunately, the current state of technology allows deploying such automated systems which can 1) administers assessments, 2) grades responses, 3) generates in-situ analytics, 4) keeps track of learning process, and 5) disseminates results to all stakeholders.

The rest of the report is organized as follows. Next section presents the related works found in the literature. Section 3 illustrates the research methodology used, with results and discussion in section 4. The authors conclude the paper in section 5, as well as set the future directions.

II. RELATED WORK

The authors build on earlier research conducted on the quizzing effect and that of the smart in-situ analytics. Albeit, we have not found—to the best of our knowledge—a study reporting a system benefiting by both domain, but studies are conducted on smart analytics or testing effect are abundant.

Macfadyen and Dawson [4] reported on the development and deployment of an automated system that mines their learning management system data to produce analytics. Their system was able to teach students’ performance, and predict at-risk learners. They suggested that such leads can possibly be used to rectify the learning problems.

Essa and Ayad [5] proposed a student success system, which subjected educational data to predictive models to identify at risk students. The system also created visualizations to easily observe the data and patterns. They postulated that such system could bridge the gap between prediction and intervention.

McDaniel et al. [6] experimented with quizzing frequency to find quizzing effect in real classrooms, contrary to earlier studies conducted in laboratory settings. They concluded that the quizzing might have improved the metacognition for the course content, resulting in significantly better performance of students in high-stake summative assessments.

In a recent meta-analysis on the subject of quizzing effect, Adesope et al. found that the retrieval practices in general increase students’ efficiency and effectiveness [7]. They suggested that the power of low-stake quizzing be exploited in K-12, as well as tertiary educational systems.

III. RESEARCH METHODOLOGY

A. The Context

The experiment was conducted at a large private sector university in Pakistan. Every year, the university intakes up to 600 students in fall and 200 students in spring, under Baccalaureate in Computer Science Program. The experiment ran for the entire duration of 6 semesters. The first two semesters were taken as the control group, where the educational process were kept in the conventional settings. The subsequent 4 semesters were treated with the novel pedagogy. The whole student intake of these semester participated in the experiment, but those who left.

The primary research problem was to reduce the number of failing and withdrawing students from the course of their studies, either at a subject’s level, or from the Baccalaureate program altogether. For the course of the study, the subject of “Functional English” (the first course offered in English as a Second Language) was selected.

B. The System

The system had three main components working in conjunction: two subsystems, namely Transparent Learning Online (TLO) and Learning Management System (LMS), and a content store accommodating data including learning content, assessments, and progress records. The purpose of the TLO was to deliver the learning content. The LMS primarily handled assessments, data generation, in-situ processing, analytic generation and smart distribution.

Albeit, the semesters ran in a conventional classroom settings, but the learning was not confined to classroom context merely, rather blended. Both subsystems were available online providing access to the content and the analytics from anywhere anytime via the Internet. However, the assessments were limited to the classroom settings only. The restricted access to the assessment module ensured that no cheating was done during assessment sessions. The schematic of the system, and the potential information flow paths are illustrated in Figure 1.

The regular teaching and assessments sessions were held in the language labs of the department. Each lab was equipped with enough computers to accommodate an entire section. The teacher, in this context, facilitated the delivery of content from the TLO and administered assessments through the LMS.

1) The TLO Subsystem

All the learning was conducted with the help of TLO subsystem. The university purchased one TLO account for each registering student beforehand. There were primarily two type of learning materials available, 1) the learning content provided by the TLO, and 2) in-house developed learning content uploaded on TLO with authoring privileges. Figure 2 depicts a few learning activities provided on TLO subsystem.

The system provided learning in the form of audio, visual, and read/write activities. Albeit, the authors wanted to implement kinesthetic, however, technological limitation did not allow the use of TLO for such learning mechanism.

2) The LMS Subsystem

The LMS subsystem was built on Moodle base by a team of in-house developers. The primary purpose of the subsystem was to administer assessment activities, grade assessment items, collect data, generate in-situ analytics, and provide customized reports to a community of users with differing
roles and views in a smart fashion. To accommodate the diversity of views, the system was built in a Model-View-Controller (MVC) fashion. Figure 3 presents a schematic view of the LMS subsystem in details.

The assessments administered via the LMS were mostly low-stake short-duration quizzes. On average, 12 such quizzes were administered during a 14-learning weeks of a 16-week semester. The remaining two weeks had high-stake mid-term and final-term exam administered in a conventional—non-LMS—fashion. Similarly, assignments and presentation activities were handled without LMS, though a submission option for assignment documents was always available. A few example of low-stake quizzes are illustrated in Figure 4.

Each student was provided with an LMS account, which they used to login to the system. A typical landing page for a student is depicted in Figure 5, where a number of possible activities can be seen. In general, the account could be used to do a variety of tasks including 1) taking assessments, 2) downloading materials, 3) submitting assignments, 4) finding announcements, and 5) to review analytics.

Along with the views available in one’s LMS account, the analytics were also available via the smart propagation subsystem. Considered to be the vital part of the system, the smart propagation subsystem sent appropriate views to all stakeholders via emails, as soon as the analytics were generated. An example of a graph aggregating progress information on a number of instruments for an entire class is shown in Figure 6.

However, the system was not limited to aggregated visuals. It allowed user to choose how coarse the information they needed, ranging from a single instrument to all instruments of an entire section, and even comparing between courses. Needless to say, the information was not arbitrarily available. The views of a teacher differed from that of a student, or from the Head of the Department (HoD). In principal, the authors identified 4 stakeholders to be kept in this closed feedback loop, namely students, teachers, parents, and heads—like head of the department, and the dean of the faculty. Each role had its own privileges, with students and parents being the most restricted, and HoD/dean being the...
most privileged. A few example views for different roles are presented in Figure 7.

C. The People

Since the observation was run for the entire duration of 6 semesters involving all the enrolling students, the number of participants was significantly large. A total of 2049 students were enrolled into the program during this period, and hence into the experiment. The fall intakes were further partitioned into 12-13 sections, and the spring were partitioned into 3-4 sections. Mostly, a section comprised 50 students, and was identified with a letter ranging from A to M. The authors considered the partitioning random because no set criteria was followed. The students were put into sections on first come (register) first serve basis. At any given time, only one section was open for registrations. A section’s registration was closed as soon as it received all 50 students registered into it, triggering the opening of the next section in line.

The schooling system in Pakistan has a considerable diversity. Multiple teaching and examination systems are running in parallel, for example, Cambridge, American, and the local. The students from all these systems are equally acceptable for admission at the institutes of tertiary education. Thus, the participants were believed to represent a diverse population, however, covering the entire spectrum.

D. The Procedure

The control group was given no treatment additional to the conventional teaching methodology of ESL in Pakistani context. For a normal semester, they had 14 teaching weeks, accompanied with two exam, 4 quizzes, 4 assignments, 1 presentation, and a few class participation activities. Most of the assessment instruments were paper based, administered, collected, graded, and notified manually. In some cases, especially where the teachers has large enrollments, the feedback was delayed for significantly longer durations.

The treatment group comprised of 4 subsequent semesters. Their assessment administration and feedback delivery mechanism were different than the control group. For a 16 week semester, they had 12 quizzes—even more in some cases—4 assignments, 2 exam, 1 presentation, and a few class participation activities. Almost all of the quizzes were administered via LMS software. The students’ attempt was instantly graded by the software, and the feedback was made available right away. Moreover, visual representation of the accumulating weekly progress of each student was shared with the teachers, with the students, as well as with the parents for facilitation. The larger enrollments, or lengthy assessments, or difficult to grade instruments, were not a problem. The entire process of producing and sharing feedback was completed within minutes, ensuring a close-knit closed feedback loop.

E. Measurement and Analysis

Each semester produced a grade distribution, though with differing number of observations, corresponding to the number of enrolled students, as detailed in Table 1. Amongst
fall semesters, the difference of the enrollment was negligible, and so was within spring semesters. But, the difference between fall and spring semesters was significantly large. Thus, the distributions were not comparable in raw number. The authors normalized the scores to averages to calculate the percentage of drop-outs and failures in each semester.

**TABLE 1. NUMBER OF STUDENTS ENROLLING IN EACH RESPECTIVE SEMESTER. PLEASE NOTE THAT THE FALL SEMESTERS USUALLY INTAKE SIGNIFICANTLY LARGE ENROLLMENT AS COMPARED TO SPRING SEMESTERS.**

<table>
<thead>
<tr>
<th>Semester</th>
<th>Number of Students</th>
</tr>
</thead>
<tbody>
<tr>
<td>S15</td>
<td>151</td>
</tr>
<tr>
<td>F15</td>
<td>452</td>
</tr>
<tr>
<td>S16</td>
<td>142</td>
</tr>
<tr>
<td>F16</td>
<td>548</td>
</tr>
<tr>
<td>S17</td>
<td>155</td>
</tr>
<tr>
<td>F17</td>
<td>601</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>2049</strong></td>
</tr>
</tbody>
</table>

The authors subjected the distributions to the visual analysis via inspection of histograms. For the comparison, each grade distribution was plotted separately, and then one joint plotting of the trend of withdrawals and failures was plotted separately.

**IV. RESULTS AND DISCUSSION**

Before the interventions into the academic process, the number of students failing the course or dropping out of the course during the semester was significantly large. Though, the actual number of enrollments in each semester differed largely (see Table 1), but the averages agreed. Around 25% of the total enrolment was dropped from the case course (English-1), with another 3% failing, as depicted by Figure 8. The agreement of numbers was not surprising to the authors, since it kept this way since long, and this very problem provided basis for this research. Furthermore, the grade distribution of both semesters from grades A to D, somehow agreed as well.

The first hi-fidelity, fully functional, pilot run was ready in Spring 2016. During the course of this semester—and in subsequent semesters—the smart analytics system kept teachers, students, and parents in an instant closed feedback loop, continually updating about the progress. The results of this semester showed a significant reduction in withdrawals. However, the percentage of the failures climbed up, as depicted in top-left of Figure 9. The actual reasons behind this fluctuation are unknown. The authors may speculate about two reasons. First, being in the pilot run, the system was not in a mature enough shape. Second, it was the first experience of teachers as well as students with the system, who may have been in a period of denial/uncertainty over accepting or rejecting the system, a dilemma already faced by the educational community [8].

In the following semesters (Fall 2016, Spring 2017, and Fall 2017), the situation improved significantly, not only for withdrawals and failures, but for the rest of grade distribution as well. The percentage of failures went down to a negligible level, and withdrawals kept around 10%. The entire grade distributions of these semesters are depicted in Figure 9.

With a regression model fit to the distributions of withdrawals (see Figure 10), each successive semester seemed to receive reduction slightly larger than a factor of three percent. The model was a good fit, with a substantial value of coefficient of determination ($R^2_w = 0.61$) [9], implying the statistical significance of the reduction rate. The rate of reduction in failures received a factor of 0.85 percent propagating into each successive semester, with a slightly moderate value of coefficient of determination for this distribution ($R^2_f = 0.23$) [9]. The intercept for both distribution were not interpreted by the authors, because the regression model was used as the test of significance, and not as a predictor model.

Albeit the coefficient of determination for failures was lesser than that of for withdrawal, i.e. $R^2_f < R^2_w$, but the authors still considered it statistically significant for a number of reasons. First, it was still in the range of moderate significance, in light of the benchmarks provided by Urbach and Ahlemann [9]. Second, the empirical evidence produced from numerical data showed a significant drop in the number of failures, receding in chronological order. Third, the reduced $R^2_f$ value may be attributed to a sudden spike (outlier) in failure rate of Spring 2016 (the semester that first received the treatment). Finally, the regression model was not fit as a predictor model. It was rather used to understand the effect already occurred in the past. Thus, the authors considered both withdrawals and failures to receive a statistically significant reduction rate.
The reader shall note that the withdrawal is a complex phenomenon, not exclusively dependent upon the learning situation and discovery of erroneous concept and its rectification [1]. Several other factors contribute to students’ withdrawing from the subject or from the program altogether. A couple of notable factors include financial problems limiting the ability to pay fee and/or inability to maintain the required attendance ratio. However, none of the enumerated reasons (or others) have been manipulated in the experiment. Their effect remains unchanged throughout six semesters.

The only reason manipulated is the timely discovery of errors via low-stake quizzing and the closed instant feedback loop. Thus, the authors attribute the reduction of withdrawals—as well as failures—primarily to the controlled variable. The in-situ generation of smart analytics, and instant delivery of feedback to all stakeholders keeps them informed of the current learning situation. Consequently, the necessary corrective measures are not delayed till the point of no recovery. It is thus advisable that such systems shall be developed and deployed to improve student retention, especially in the context of large enrollments and/or heavy teaching load.

V. IMPLICATIONS

Considering the importance of keeping the students, parents and teachers informed of the progress during the semester, the system provides succinct information. The sharing of results with all the stakeholders provide a number of opportunities.

First, the students have the right information at the right time. They can go back to the problematic topic—which happens to be the previous week—and learn it afresh again. It shall reduce the errors in that very topic, as well as limit the chances of its ripples extending in the forthcoming. The result is a more confident student less likely to withdraw from the subject.

Second, the teachers have the right information at the right time. When they receive the results instantly (especially smart analytics), they are in a better position to decide whether they need to move forward, or stay for a while to cover the older topics again. They can also arrange extra tutorials for the students lacking behind.

Third, the parents—who have a strong influence in the context of Pakistani educational system—are informed of the situation of their children. They can counsel the student to focus on their studies.

Finally, a student counselor can intervene into the situation if the student is not able to recover up. Additionally, if a class is, in general, lacking behind, the head of the department can discuss it with the teacher or take some corrective measures.

VI. CONCLUSIONS AND FUTURE WORK

The purpose of the study was to investigate the effect of the novel pedagogical model—combining frequent low-stake formative assessment with smart analytics and instant feedback—on the retention of students of Baccalaureate of Computer Science. The impact of the proposed model was examined through academic achievement of the students. In general, the authors found a statistically significant decrement trend in the rate of withdrawals and failures in the case subject, namely “Functional English”.

In the future, the authors want to extend the system to accommodate the learning needs and behaviors of an individual learner via adaptive hypermedia technology. The expected benefits of this extension are better performance on the entire grade spectrum.

ACKNOWLEDGEMENTS

We highly appreciate the participation and support provided by the members of the faculty and students for this research. We are also thankful to the management and the Dean of the faculty for approving arrangement of conducting in-situ experiments.

REFERENCES


