Are Working Habits Different Between Well-Performing and at-Risk Students in Online Project-Based Courses?

Mingxiao An
Carnegie Mellon University
Pittsburgh, PA, USA
mingxiaa@andrew.cmu.edu

Hongyi Zhang
Carnegie Mellon University
Pittsburgh, PA, USA
hongyiz2@andrew.cmu.edu

Jaromir Savelka
Carnegie Mellon University
Pittsburgh, PA, USA
jsavelka@andrew.cmu.edu

Shijie Zhu
Carnegie Mellon University
Pittsburgh, PA, USA
shijiez@andrew.cmu.edu

Christopher Bogart
Carnegie Mellon University
Pittsburgh, PA, USA
cbogart@cmu.edu

Majd Sakr
Carnegie Mellon University
Pittsburgh, PA, USA
msakr@cs.cmu.edu

ABSTRACT

We analyze differences in working habits between well-performing and at-risk students using highly-granular data collected from two semesters of an online project-based, upper-level course on cloud computing at a US institution of higher education. Such differentiating metrics may provide deeper insights than interim grades, which are oftentimes the only quantifiable data that is captured and available to an instructor as a proxy for students' learning. Interim grades provide little insight into students' broader work habits and may mask unsustainable learning strategies that result in shallow learning or quickly-forgotten skills/knowledge. The adoption of technology-enhanced learning tools for course delivery, automatic feedback, and grading enable data-informed insight and reflection into students' working habits. This data could allow the detection of early signs of under-prepared students or students in crisis.

We empirically assess what working habits, if any, differ among well-performing and at-risk students. From clickstream and other activity data, we derive 22 metrics such as time spent reading project write-ups, timing of starting and finishing work, or break-taking. We also calculate two measures of consistency of each metric measured by a coefficient of variance and a variance of ranking over the semester as well as outlier behavior of a student. Using Z-test and Kolmogorov-Smirnov test, we confirm differences in multiple behavior patterns. Notably, our data suggest that well-performing students start and finish working on a project earlier than at-risk students but they also tend to have fewer submissions which indicate they are more thoughtful about feedback.

CCS CONCEPTS

• **Social and professional topics** → **Student assessment.**

KEYWORDS

online education, student performance, educational data mining, student behavior, work habits

This work is licensed under a Creative Commons Attribution International 4.0 License.

ITiCSE 2021, June 26–July 1, 2021, Virtual Event, Germany.

© 2021 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-8214-4/21/06.

<https://doi.org/10.1145/3430665.3456320>

ACM Reference Format:

Mingxiao An, Hongyi Zhang, Jaromir Savelka, Shijie Zhu, Christopher Bogart, and Majd Sakr. 2021. Are Working Habits Different Between Well-Performing and at-Risk Students in Online Project-Based Courses?. In *26th ACM Conference on Innovation and Technology in Computer Science Education V. 1 (ITiCSE 2021)*, June 26–July 1, 2021, Virtual Event, Germany. ACM, New York, NY, USA, 7 pages. <https://doi.org/10.1145/3430665.3456320>

1 INTRODUCTION

This work explores differences in working habits between well-performing and at-risk students. We present an in-depth analysis of data that capture how students work on projects from the moment the project write-up becomes available online via a learning management system (LMS) to their last submission. We analyze data from eleven weekly projects that together form the bulk of the work of an online project-based course on cloud computing.

Student work in CS/IT education (e.g. algorithms, databases, and networks) is often “invisible” to instructors until very far along, or even complete [9] making it difficult for faculty to provide early feedback. This invisibility is even more pronounced in settings where instructor contact with students is limited because of large class sizes, online teaching, or limited budgets to hire TAs. Data-driven support of instructor oversight of student work [25] requires measuring and displaying what students are doing, and communicating that in a meaningful way. While data dashboards have become quite common, research has shown that it is difficult to provide useful, actionable data about student behavior to instructors during the course, beyond the gradebook data they typically already have [12, 20].

Our larger research aim is to improve instructors' ability to intervene to improve student outcomes, by helping them oversee students' work patterns and habits, not just grades; offering predictions and metrics to augment instructors' intuitions about effective online learning behaviors, and to suggest interventions likely to improve the outcomes for students who appear to be at risk. In this paper we work towards that goal by evaluating a pool of metrics, derived from raw log data from a cloud computing course, of student behaviors that we hypothesize to be evidence of good work habits. The assessment of those metrics could enable instructors to build a better intervention strategy.

We briefly outline related work in Section 2. In Section 3 we describe the cloud computing course and the way students interact with our infrastructure. We describe data we collect in Section 4. In

Section 5 we explain the method of analyzing the data. We present our results in Section 6. Section 7 provides an in-depth discussion of the findings. We conclude the paper (Section 8) and suggest several directions for future work (Section 9).

2 RELATED WORK

There is an abundance of research focused on finding predictors for students' performance. Several studies [1, 5] attribute students' success to their abilities and mental models measured by tests and quizzes. Seidman [22] uncovers the disruptive behavior in class that could impair learning. However, in an online learning environment, such face-to-face interaction data are not available. Based on the data collected from several MOOCs, Qiu et al. [19] find that learning behavior factors, such as the number of chapters a student browses, are the most useful predictors of student performance. In [7] the researchers present an approach to detect patterns in student submissions that can predict final performance with 79% accuracy using only the first programming assignment.

Several studies investigate how behavior relates to students' final performance. Some studies [4, 16, 17] report a positive correlation between time spent and final performance. Yet others [6, 18, 19, 21, 24] suggest that effort or activity level is a weak predictor of student performance. Some [14, 23] also suggest that total studying time does not predict final grade, they point out that several factors, such as login frequency and (ir)regularity of learning interval in the LMS, had a significant correlation with final grades. Finally, Masui et al. [16] report that the relationship between effort and performance is complicated and depends on context.

Wolff et al. [27] use clickstream data to predict at-risk students, but does not focus on explanation or gaining actionable insights for instructors. Several studies [10, 13, 15, 26] identify beneficial as well as problematic patterns in learning behavior of high and low performing students. Chen et al. [3] suggest that taking context into account (when, where, and how) can lead to better understanding of students' learning and more actionable insights for instructors.

Goldstein et al. [8] investigate how student behavior and its consistency are related to the final performance and analyze which kinds of behavior factors are most relevant. They conclude that the consistency of student behavior might be more predictive of performance than behavior itself.

In this work, we compare the work habits of well-performing students to the work habits of at-risk students. We use statistical analysis to determine which differences are significant. Finally, we provide a detailed discussion of the findings focusing on actionable insights that could lead to successful instructor interventions.

3 BACKGROUND

In this study we work with data collected from two consecutive runs of a semester-long online project-based course on cloud computing offered to students of a single US-based University at several global campuses in Spring and Fall 2020. The course is open to undergraduate and graduate students. The main bulk of work, which all of the students are required to complete, consists of eleven programming projects. The graduate students also need to participate in a team project. In this study, we focus on students' engagement with the course materials related to the eleven programming projects.

Figure 1: The distribution of the score averages from the top ten most successful projects per student. The graph on the left shows the distribution for the Spring 2020 semester while the graph on the right presents data from the Fall 2020. The students above the green line (0.95) are considered the well-performing students while those below the red line (0.85) are presumed to be the at-risk students.

Since this is an online course, there are no lectures to attend, although there are multiple other activities the students are involved in (e.g., conceptual topics, project write-up section gateway quizzes, collaborative programming exercises). Students interact with the teaching staff and with other students on an educational question-and-answer forum, Piazza.¹ The course has been offered 14 times before in the same format.

We use our own platform, TheProject.Zone, to deliver the programming projects. Students submit solutions using provided submitter programs to our own auto grading service at TheProject.Zone. The service evaluates the solution and presents the score and feedback to the student. Students can use the feedback to iterate on the solution and re-submit the corrected/improved version (there are no limits on the number of submissions). The evaluation may involve anything from validating a student program's output, testing its runtime properties, or interacting with the web services it provides. In order to complete the project, the students need to read and understand topic-specific primers, then solve the tasks defined in a project write-up. The primers are meant to ensure students have the knowledge and skills required to engage with the project tasks in a meaningful way.

4 DATA

The data collected and reported on in this study are from undergraduate and graduate students working on the individual projects only. Data were collected from 170 students enrolled in the course in the Spring 2020 semester and 69 students from the Fall 2020 class. The data come from the students' interaction with TheProject.Zone, captured and preserved by our logging and monitoring pipeline. In this paper we use clickstream data including *navigations* to content pages, *heartbeats* (generated every 5 minutes when a student's browser is open on a content page), assignments *submissions*, and *quiz attempts* for in-line project write-up gating quizzes.

We transform the raw data related to students' behavior into factors that are easily understandable by instructors. For example, we utilize heartbeat to compute factors such as students taking short versus long breaks as well as their frequency or the time of first/last

¹<https://piazza.com/>

navigation to the module write-up. Similarly, we use submission data to represent students' improvement rate or time spent on improving their solutions. The full list of 22 student behavior factors is available in Table 1.

These 22 raw factors are the basis for computation of additional factors, which include 22 absolute consistency, 22 relative consistency, and 22 outlier factors (see Section 5.1 for details). Ultimately, we are interested if and how any of the 88 factors differ between well-performing students and students that may struggle (at-risk students). While the ultimate goal is to identify which work habits are correlated with the best learning outcomes, we rely on grades as a proxy for learning. Per our working definition well-performing students are those who obtained at least 95% average score from their projects while the at-risk students are those with the average of 85% or less. The threshold of 85% may appear to be set quite high for a student to be deemed at-risk. Note, that being at-risk does not automatically imply a student struggles. It merely points to the fact that the student might be at-risk of struggling at certain point of the course. Also note, that the least successful project does not count towards the grade and hence, in this study we do not consider such projects either (i.e., for each student their least successful project is omitted from the analysis). It is also removed from the dataset if the student intentionally give up a module. The distributions of the score averages for both semesters, including the cut-off lines for the well-performing and at-risk students, are shown in Figure 1.

5 METHOD

In this section, we first introduce the various types of behavior features we extract and generate from the student online activity logs. Then, we explain how we use Spearman's rank correlation coefficient to focus on those features that correlate with student performance measured in terms of the grades they obtained. Finally, we describe how we evaluate the selected subset of features using Z-test and Kolmogorov-Smirnov test to determine the differences between the behavior of well-performing and at-risk students.

5.1 Student Behavior Features

5.1.1 Averaged Factors (AvF). Raw factors are project-level metrics derived from the student activity logs. They are meant to be intuitive to instructors, e.g., starting earliness, quiz correctness, or break frequency. Altogether, there are 22 such factors (Table 1). For each student, we compute the factor average over their ten most successful projects. The averaged factor is defined as:

$$f_{avg}^{(i)} = \frac{1}{|\mathcal{P}|} \sum_{p \in \mathcal{P}} f_p^{(i)}, \text{ where}$$

p an individual project in $\mathcal{P}$
 $f_p^{(i)}$ the i -th factor in project p

5.1.2 Consistencies. We also measure the consistencies of raw factors across projects to reflect how much the students adhere to their established working habits over the duration of the course. The *absolute consistency (AbC)* of a raw factor is defined as the coefficient of variance across all projects, i.e.:

$$f_{abs}^{(i)} = \frac{1}{f_{avg}^{(i)}} \sqrt{\frac{\sum_{p \in \mathcal{P}} (f_p^{(i)} - f_{avg}^{(i)})^2}{|\mathcal{P}| - 1}}$$

This measure describes how consistent is a student with respect to a particular work habit when compared to their behavior on other projects (e.g., Do they always start their work early?).

Since the difficulty of the projects as well as their context varies (e.g., holidays) we compute the *relative consistency (ReC)*. This measure describes how consistent is a student with respect to a particular work habit when compared to their peers (e.g., Are they always among the first to submit their work?). For example, most students would have a longer total working time on a difficult module. For each factor, the relative consistency is measured as the standard deviation on the percentile factor ranking for each module:

$$f_{rel}^{(i)} = \sqrt{\frac{\sum_{p \in \mathcal{P}} (r_p^{(i)} - \bar{r}^{(i)})^2}{|\mathcal{P}| - 1}}, \text{ where}$$

$$r_p^{(i)} = \frac{1}{N_p} \text{Rank}(f_p^{(i)});$$

$$\bar{r}^{(i)} = \frac{1}{|\mathcal{P}|} \sum_{p \in \mathcal{P}} r_p^{(i)};$$

N_p the number of students participating in project p .

5.1.3 Outlier Factors (LOF). Atypical behavior can be a potential indicator of at-risk students, which would in turn be a very informative and helpful to an instructor. We extract local outlier factors [2] for each raw factor to detect atypical behaviors. Given a distance function $d(\cdot, \cdot)$, the local reachability density distance for student s with maximum K nearest neighbours is defined as:

$$lrd^{(i)}(s) = \frac{K}{\sum_{k=1}^K \max \left\{ \frac{d(f_{con}^{(i)}(s), f_{con}^{(i)}(N_k(s)))}{d(f_{con}^{(i)}(N_k(s)), f_{con}^{(i)}(N_K(N_k(s))))} \right\}}, \text{ where}$$

$f_{con}^{(i)}(s)$ the concatenation of $f_p^{(i)}(s)$ for each project p of student s and factor i ;
 $N_k(s)$ k -th nearest neighbour of student s .

The local outlier factor of the student is then defined by the average lrd of the K neighbours over its own lrd , i.e.:

$$f_{lof}^{(i)}(s) = \frac{K \cdot lrd^{(i)}(s)}{\sum_{k=1}^K lrd^{(i)}(N_k(s))}.$$

The algorithm predicts outliers from similar students without an explicit definition of data clusters. That is the main reason why we prefer local outlier factor over k-means or DBSCAN. As a distance metric, we use the Euclidean distance.

5.2 Recognizing the Most Important Features

To find out which work habits are important in student learning, we use Spearman's rank correlation with the performance as the key metric. Spearman's correlation is the statistical correlation on the ranking of two sets of data. We prefer this method because ranking enforces uniform distribution of data, which reduces sensitivity to extreme data points. The r_s statistic is computed as follows:

$$r_s = \rho_{rg_X, rg_Y} = \frac{\text{cov}(rg_X, rg_Y)}{\sigma_{rg_X} \sigma_{rg_Y}}, \text{ where}$$

- ρ the usual Pearson correlation coefficient applied to the rank variables (RVs)
- $\text{cov}(rg_X, rg_Y)$ the covariance of the RVs
- σ_{rg_X} and σ_{rg_Y} the standard deviations of the RVs

We use the p-value of the r_s statistic (defined above) from the Spring 2020 semester to determine which work habits correlate with students' performance as measured by their grades. Since the p-value is a measure of how likely it is that the observed correlation is due to chance, we adopt an assumption that any work habit with p-value < 0.01 correlates with the performance. Note, that we did not employ Holm-Bonferroni method [11] as a protection against Type I errors. As our sample size is relatively small, using the correction would almost certainly increase the danger of Type II errors to unacceptable levels. To increase the persuasiveness of our analysis, we plan to utilize the Holm-Bonferroni method in the future when more data is available.

5.3 Evaluation of Differences Between well-performing and at-Risk Students

After identifying which of the work habits appear to correlate with performance, we investigate if and how these habits differ between well-performing and at-risk students. We first conduct a standard Z-test to test the difference between means of the work habit distributions associated with each group. Additionally, we use the Kolmogorov–Smirnov test (KS test) to test the largest difference of the cumulative distributions:

$$D_n = \sup_x |F_n(x) - F(x)|, \text{ where}$$

- $\sup_x$ the supremum of the set of distances
- $F_n(x)$ $\frac{1}{n} \sum_{i=1}^n I_{[-\infty, x]}(X_i)$
- $I_{[-\infty, x]}(X_i)$ the indicator function, equal to 1 if $X_i \leq x$ and equal to 0 otherwise

Both tests assume the same distributions as the null hypothesis. Lower p-values then stand for stronger rejection, proving the distributions are different. We perform the tests on both of the semesters where the Fall 2020 is understood as a validation (since the initial filtering was performed on Spring 2020 semester). We do not use Holm-Bonferroni correction for reasons explained above.

6 RESULTS

Table 1 shows the results of the analysis focused on discovering which work habits correlate with the grades. Each work habit entry is associated with a bar chart where each bar represents the strength of the correlation between the performance (grades) and a work habit as a raw factor, absolute and relative consistency, as well as the local outlier factor. Overall, we identified eighteen features with p-value < 0.01. These features are listed in Table 2. While some of the work habits make it through the filter in multiple modalities (e.g., Active Working Time), most of them appear only under single transformation (typically as a raw factor or absolute consistency).

The results of the analysis focused on the evaluation of differences between well-performing and at-risk students are reported

Figure 2: Distributions of First Submission in both semesters. Higher values mean later submission. Well-performing students tend to make the the first submission earlier. The y-axis stands for standard deviation.

in Table 2. The p-values of the statistics generated by the two tests for the Spring 2020 are very low (all < 0.05) for all the 18 work habits we focus on at this stage of the study. This is to be expected as the pre-selection of the features was performed on the same semester. Hence, we already had very strong evidence that there is a correlation (beyond chance) between these work habits and student performance. Then, in turn it should not be surprising that we also detect differences in those habits between the well-performing and at-risk students. This is precisely why it is important to validate the results on a different semester (Fall 2020). There we see that by far not all of the p-values fall below the 0.05 threshold. At the same time, there is a number of work habits where the differences between the two groups are confirmed.

7 DISCUSSION

In this section we provide our analysis and interpretation of the results presented in Section 6. The work habits that we discuss can be roughly categorized into four higher-level groups—those related to the time students start their work on a particular project (Subsection 7.1), to the time students spend working on a project (Subsection 7.2), to the breaks students take during the work (Subsection 7.3), and to the time students finish their work (Subsection 7.4).

7.1 Starting Time

The data from both semesters strongly suggest that well-performing students start to work on the projects earlier than at-risk students. While this behavior is observed at multiple elements related to their work, such as the first time they read the project write-up (First Navigation, First Delayed Navigation), or the first time they take a gating quiz (First Quiz), it is especially pronounced in the time they make their first submission of a task solution (First Submission) as shown in Figure 2. The fact that they make an early submission is a very strong indicator that they start serious work on the project early; as opposed to just casually skimming the project write-up.

Interestingly, we were not able to confirm that the well-performing students are consistent when it comes to times of their first submission, navigation, or taking a gating quiz. This is somewhat surprising as it appears that they indeed are quite consistent when it comes to the times of their last submission (see Subsection 7.4). A possible cause of students being able to make the first submission earlier might be their stronger prior background or an actual habit of getting their hands dirty rather soon in order to get a better idea of how complex a project is.

Table 1: Student behavior factors derived from activity logs. Bars show Spearman’s correlation with project grades, measured by Blue: raw factor Orange: absolute (within-student) consistency Green: Consistency relative to other students Red: Outlier status. Note that right-facing bars are high for raw factors, but consistency is a negative measure. For example the first row shows that well-performing students are likely to take short breaks (blue) consistently (orange).

Table 2: KS-test and Z-test results on the distribution difference of well-performing and at-risk students. P-value with * is lower than 0.05; p-value with ** is lower than 0.01.

Measure	Factor Name	S20				F20			
		KS stat	KS p-value	Z-test stat	Z-test p-value	KS stat	KS p-value	Z-test stat	Z-test p-value
Averaged Factor	First Navigation	0.2877	$3.00 \times 10^{-2*}$	-3.5872	$3.34 \times 10^{-4**}$	0.5368	$9.57 \times 10^{-3**}$	-2.4311	$1.51 \times 10^{-2*}$
Averaged Factor	First Delayed Navigation	0.4116	$4.62 \times 10^{-4**}$	-4.0370	$5.41 \times 10^{-5**}$	0.5368	$9.57 \times 10^{-3**}$	-2.5010	$1.24 \times 10^{-2*}$
Averaged Factor	First Submission	0.5406	$7.68 \times 10^{-7**}$	-6.3810	$1.76 \times 10^{-10**}$	0.6421	$8.86 \times 10^{-4**}$	-5.0611	$4.17 \times 10^{-7**}$
Averaged Factor	First Quiz	0.3790	$1.56 \times 10^{-3**}$	-3.5166	$4.37 \times 10^{-4**}$	0.4456	$4.83 \times 10^{-2*}$	-2.4232	$1.54 \times 10^{-2*}$
Averaged Factor	Pre-Submission Time	0.4389	$1.25 \times 10^{-4**}$	-5.3066	$1.12 \times 10^{-7**}$	0.6175	$1.55 \times 10^{-3**}$	-4.1400	$3.47 \times 10^{-5**}$
Averaged Factor	Average Task Submissions	0.3042	$1.84 \times 10^{-2*}$	-2.9632	$3.04 \times 10^{-3**}$	0.4596	$3.78 \times 10^{-2*}$	-2.1799	$2.93 \times 10^{-2*}$
Averaged Factor	Active Working Time	0.4628	$4.81 \times 10^{-5**}$	3.0355	$2.40 \times 10^{-3**}$	0.2281	6.80×10^{-1}	1.0268	3.04×10^{-1}
Abs-Consistency	Active Working Time	0.5485	$5.41 \times 10^{-7**}$	-5.7959	$6.80 \times 10^{-9**}$	0.5228	$1.27 \times 10^{-2*}$	-2.6479	$8.10 \times 10^{-3**}$
Outlier Factor	Active Working Time	0.3433	$5.94 \times 10^{-3**}$	-3.6090	$3.07 \times 10^{-4**}$	0.3018	3.40×10^{-1}	-1.5846	1.13×10^{-1}
Averaged Factor	Very Short Breaks	0.4715	$3.18 \times 10^{-5**}$	4.5377	$5.69 \times 10^{-6**}$	0.3509	1.97×10^{-1}	1.2233	2.21×10^{-1}
Averaged Factor	Short Breaks	0.4894	$1.25 \times 10^{-5**}$	4.3440	$1.40 \times 10^{-5**}$	0.2982	3.60×10^{-1}	0.8349	4.04×10^{-1}
Averaged Factor	Long Breaks	0.2999	$2.32 \times 10^{-2*}$	2.8624	$4.21 \times 10^{-3**}$	0.2351	6.37×10^{-1}	0.4699	6.38×10^{-1}
Abs-Consistency	Very Short Breaks	0.4711	$3.23 \times 10^{-5**}$	-4.9684	$6.75 \times 10^{-7**}$	0.4737	$2.88 \times 10^{-2*}$	-1.8554	6.35×10^{-2}
Abs-Consistency	Short Breaks	0.3429	$6.01 \times 10^{-3**}$	-3.5175	$4.36 \times 10^{-4**}$	0.4281	6.60×10^{-2}	-1.7734	7.62×10^{-2}
Abs-Consistency	Long Breaks	0.3177	$1.37 \times 10^{-2*}$	-3.2939	$9.88 \times 10^{-4**}$	0.3228	2.76×10^{-1}	-1.1190	2.63×10^{-1}
Averaged Factor	Last Submission	0.5997	$2.30 \times 10^{-8**}$	5.4495	$5.05 \times 10^{-8**}$	0.7754	$1.84 \times 10^{-5**}$	5.3651	$8.09 \times 10^{-8**}$
Abs-Consistency	Last Submission	0.5828	$6.70 \times 10^{-8**}$	-6.1465	$7.92 \times 10^{-10**}$	0.7368	$5.53 \times 10^{-5**}$	-4.6638	$3.10 \times 10^{-6**}$
Rel-Consistency	Last Submission	0.4789	$2.15 \times 10^{-5**}$	4.4542	$8.42 \times 10^{-6**}$	0.5895	$3.13 \times 10^{-3**}$	4.9134	$8.95 \times 10^{-7**}$

7.2 Working Time

Well-performing students tend to be more consistent in the amount of time they invest (Figure 3a). A similar finding is reported in [8]. Well-performing students also tend to make fewer submissions than at-risk students (Figure 3b). This suggests that well-performing students use feedback more effectively and thus do not have to make

as many submissions to get to the correct solution. Well-performing students also appear to spend less time working on their projects before attempting their first submission (Pre-Submission Time).

While the Spring 2020 semester data suggest that well-performing students spend more time working on the project (Active Working Time), the data from Fall 2020 semester do not confirm this

Figure 3: Distributions of Active Working Time (AbC) and Average Task Submissions (AvF). Well-performing students are more consistent in the time they spend per project (left) and tend to make less submissions per task (right).

Figure 4: Distributions of Very Short Breaks (AbC). High values mean more variability. Well-performing students are more consistent in taking very short breaks.

pattern. However, our data strongly suggest that from an instructor's perspective the consistency in the amount of working time might be a more important distinction than its raw value. Thus, having two students with comparable performance one could tentatively expect that the one who invests comparable amounts of time in each project will likely be better off in the end than the one whose time investment wildly varies across different projects.

7.3 Frequency of breaks

The largest group of working habits included in the analysis corresponds to those related to taking breaks from work. Interestingly, we were not able to confirm a very likely difference in these working habits, with the possible exception of the consistency in taking very short breaks (Figure 4). Well-performing students may tend to take more such breaks, but it appears that it is more important to stick to an established routine than what this routine actually is.

7.4 Finishing Time

Well-performing students tend to make their last submission earlier than at-risk students, and they appear to do so consistently. This may point to the well-performing students being disciplined at sticking to their study plan. This is because consistency in the last submission time would require a student to start, work, and (perhaps) take breaks in a very consistent manner. This discipline may also help well-performing students be less affected by varying difficulty of the individual projects.

Figure 5a clearly shows that the well-performing students finish their work earlier. This could be due to them starting earlier or already having a strong background that helps them tackle the task. Figure 5b shows that they are much more consistent when it comes to the time they make their last submission, suggesting that the earlier submissions when compared to the at-risk students are

Figure 5: Distribution comparisons of Last Submission (AvF - right, AbC - left). High values mean finishing early (left), and more variability in finishing time (right). Well-performing students finish earlier at consistent times.

rather due to how the well-performing students organize their work than due to having a stronger prior background (although, that may certainly be a factor). It is also possible that at-risk students are working up until the deadline and run out of time.

8 CONCLUSIONS

In this study we analyzed data from two consecutive semesters to determine if there are work habits that differ between well-performing and at-risk students in an online project-based course. Our data appear to suggest that well-performing students start working on a project earlier, including making an early first submission. This may provide them with an early insight into the subject matter of the project as well as its complexity. Although some instructors might promote spending more time on projects and taking many short breaks, we did not find evidence that well-performing students followed these practices. On the other hand, it appears clear that establishing a consistent routine of (not) taking breaks is important. Additionally, the data from the two semesters suggest that taking more time thinking about feedback is beneficial as opposed to fast iteration on the solution without spending proper time on improvement. Finally, the well-performing students finish their work earlier at consistent times. Overall, it appears the well-performing students are more strategic about completing the projects. They make sure they understand the nature of the project early; they are consistent in their behavior when it comes to taking breaks; and they do not rush to make a corrected submission unless they invest proper time in understanding the feedback. All this results in timely submissions well before the project deadline.

9 FUTURE WORK

In future work, we plan to study the effects of students' prior knowledge and capabilities, by using pre-tests and post-tests, rather than grades only, to confirm whether learning behavior correlates with student performance. It also remains to be studied whether there are clusters of different successful work habits suggesting different useful strategies. Understanding strategy diversity will help us distinguish outliers from truly at-risk students, and study what effects different instructor interventions have on grades as well as on working habits. Finally, we are interested in the causes of differences we detected between Spring and Fall semesters.

ACKNOWLEDGMENTS

This work was supported by Microsoft Azure Educator Grant.

REFERENCES

- [1] Jens Bennedsen and Michael E Caspersen. 2006. Abstraction ability as an indicator of success for learning object-oriented programming? *ACM Sigcse Bulletin* 38, 2 (2006), 39–43.
- [2] Markus M Breunig, Hans-Peter Kriegel, Raymond T Ng, and Jörg Sander. 2000. LOF: identifying density-based local outliers. In *Proceedings of the 2000 ACM SIGMOD international conference on Management of data*. 93–104.
- [3] Zhongzhou Chen, Mengyu Xu, Geoffrey Garrido, and Matthew W Guthrie. 2020. Relationship between students' online learning behavior and course performance: What contextual information matters? *Physical Review Physics Education Research* 16, 1 (2020), 010138.
- [4] Damian S Damianov, Lori Kupczynski, Pablo Calafiore, Ekaterina P Damianova, Gökçe Soydemir, and Edgar Gonzalez. 2009. Time spent online and student performance in online business courses: A multinomial logit analysis. *Journal of Economics and Finance Education* 8, 2 (2009), 11–22.
- [5] William EJ Doane. 2008. *Predicting student performance in introductory computer programming courses*. State University of New York at Albany.
- [6] Rosalina Rebuscas Estacio and Rodolfo Callanta Raga Jr. 2017. Analyzing students online learning behavior in blended courses using Moodle. *Asian Association of Open Universities Journal* (2017).
- [7] Ge Gao, Samiha Marwan, and Thomas W Price. 2021. Early Performance Prediction using Interpretable Patterns in Programming Process Data. In *Proceedings of the 52nd ACM Technical Symposium on Computer Science Education*. 342–348.
- [8] Seth Copen Goldstein, Hongyi Zhang, Majd Sakr, Haokang An, and Cameron Dashti. 2019. Understanding how work habits influence student performance. In *Proceedings of the 2019 ACM Conference on Innovation and Technology in Computer Science Education*. 154–160.
- [9] Mark Guzdial and Briana Morrison. 2016. Growing computer science education into a STEM education discipline. *Commun. ACM* 59, 11 (2016), 31–33.
- [10] Leslie Harvey and Ashish Aggarwal. 2020. How do Quiz and Homework Submission Times Affect Students' Performance in a Flipped CS1 Class?. In *Proceedings of the 2020 ACM Conference on International Computing Education Research*. 314–314.
- [11] Sture Holm. 1979. A simple sequentially rejective multiple test procedure. *Scandinavian journal of statistics* (1979), 65–70.
- [12] Kenneth Holstein, Bruce M McLaren, and Vincent Aleven. 2017. Intelligent tutors as teachers' aides: exploring teacher needs for real-time analytics in blended classrooms. In *Proceedings of the seventh international learning analytics & knowledge conference*. 257–266.
- [13] Wu-Yuin Hwang, Rustam Shadiev, Chin-Yu Wang, and Zhi-Hua Huang. 2012. A pilot study of cooperative programming learning behavior and its relationship with students' learning performance. *Computers & education* 58, 4 (2012), 1267–1281.
- [14] Il-Hyun Jo, Taeho Yu, Hyeyun Lee, and Yeonjoo Kim. 2015. Relations between student online learning behavior and academic achievement in higher education: A learning analytics approach. In *Emerging issues in smart learning*. Springer, 275–287.
- [15] John S Kinnebrew, Kirk M Loretz, and Gautam Biswas. 2013. A contextualized, differential sequence mining method to derive students' learning behavior patterns. *Journal of Educational Data Mining* 5, 1 (2013), 190–219.
- [16] Chris Masui, Jan Broeckmans, Sarah Doumen, Anne Groenen, and Geert Molenberghs. 2014. Do diligent students perform better? Complex relations between student and course characteristics, study time, and academic performance in higher education. *Studies in Higher Education* 39, 4 (2014), 621–643.
- [17] James W Michaels and Terance D Miethe. 1989. Academic effort and college grades. *Social Forces* 68, 1 (1989), 309–319.
- [18] E Ashby Plant, K Anders Ericsson, Len Hill, and Kia Asberg. 2005. Why study time does not predict grade point average across college students: Implications of deliberate practice for academic performance. *Contemporary Educational Psychology* 30, 1 (2005), 96–116.
- [19] Jiezhong Qiu, Jie Tang, Tracy Xiao Liu, Jie Gong, Chenhui Zhang, Qian Zhang, and Yufei Xue. 2016. Modeling and predicting learning behavior in MOOCs. In *Proceedings of the ninth ACM international conference on web search and data mining*. 93–102.
- [20] María Jesús Rodríguez-Triana, Luis P Prieto, Alejandra Martínez-Monés, Juan I Asensio-Pérez, and Yannis Dimitriadis. 2018. The teacher in the loop: Customizing multimodal learning analytics for blended learning. In *Proceedings of the 8th international conference on learning analytics and knowledge*. 417–426.
- [21] Howard Schuman, Edward Walsh, Camille Olson, and Barbara Etheridge. 1985. Effort and reward: The assumption that college grades are affected by quantity of study. *Social Forces* 63, 4 (1985), 945–966.
- [22] Alan Seidman. 2005. The learning killer: Disruptive student behavior in the classroom. *Reading improvement* 42, 1 (2005), 40–47.
- [23] Clifford A Shaffer and Stephen H Edwards. 2011. Scheduling and student performance. In *Proceedings of the 16th annual joint conference on innovation and technology in computer science education*. 331–331.
- [24] MM Van den Hurk, HAP Wolfhagen, DHJM Dolmans, and CPM Van der Vleuten. 1998. The relation between time spent on individual study and academic achievement in a problem-based curriculum. *Advances in Health Sciences Education* 3, 1 (1998), 43–49.
- [25] Janet Walkoe, Miriam Sherin, and Andrew Elby. 2020. Video tagging as a window into teacher noticing. *Journal of Mathematics Teacher Education* 23, 4 (2020), 385–405.
- [26] Salla Willman, Rolf Lindén, Erkki Kaila, Teemu Rajala, Mikko-Jussi Laakso, and Tapio Salakoski. 2015. On study habits on an introductory course on programming. *Computer Science Education* 25, 3 (2015), 276–291.
- [27] 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. In *Proceedings of the third international conference on learning analytics and knowledge*. 145–149.