MULTI-FACTORIAL PATTERNS OF ONLINE HOMEWORK USAGE IN ENGINEERING: A PILOT STUDY

Agnes d’Entremont¹, Jonathan Verrett², ShunFu Hu³, Juan Abello¹, Negar M. Harandi⁴, Tyler Leoncio¹, Warren Chan Wan Fong⁵

Departments of ¹Mechanical Engineering, ²Chemical and Biological Engineering, ³Educational and Counselling Psychology and Special Education, and ⁴Electrical and Computer Engineering, University of British Columbia

adentremont@mech.ubc.ca, jonathan.verrett@ubc.ca, abello@mech.ubc.ca, negarm@ece.ubc.ca

Abstract – WeBWorK online homework usage data for a second-year, 130-student mechanical engineering course was analyzed using latent profile analysis (LPA) to identify student usage patterns and their relation to tests/exams grades. Ten WeBWorK usage variables were used by LPA to identify three distinct student sub-groups having particular usage patterns. The resulting three sub-groups were found to have statistically significant differences in tests/exam grades. Lower grades corresponded to fewer WeBWorK sessions and questions attempted, with a higher number of attempts and questions attempted per session; lower grades also corresponded to lower collaboration metrics and later first submissions of correct answers. These results might be used by instructors to inform and encourage online homework usage practices that are related to higher grades.

Keywords: Online homework, WeBWorK, homework submission patterns, learning outcomes, latent profile analysis

1. INTRODUCTION

Online homework systems have benefits to students and instructors. Through online homework systems, students get immediate feedback and can reattempt problems to improve understanding. Faculty benefit from a reduced marking burden, which is especially beneficial in large classes. The homework problems can also be set up to randomize problem variables so that each problem set is unique for each student.

WeBWorK is an online homework system used by more than 1300 institutions worldwide that allows instructors to both create custom problems and draw from an open problem library with over 35,000 problems [1]. This library is comprised of material from multiple institutions and covers a variety of subjects [1].

Limited work has been done on students’ use of online homework systems. Existing work suggests that more time spent on an assignment is associated with better assignment grades [2], and that earlier homework problem attempts during the assignment period [3], [4] and collaboration between students [4] are associated with better exam grades. Additionally, some studies have found that the use of online homework systems had a positive impact on student grades overall [5].

Some research has been done exploring student attitudes towards online homework systems. Prior studies suggest that students view online homework “favorably” [5] and found it to be an “effective method of study” [6]. Studies also found that online homework motivated students towards “more consistent study” and resulted in higher homework completion rates [5].

Analyzing student online homework usage patterns and their relationships to academic performance may provide some additional insight into how students best learn. At present it is unclear whether additional factors (or combined patterns of factors) might also be associated with particular student academic outcomes.

In this study, we explore whether student WeBWorK usage patterns can be identified over multiple system data metrics, and whether these patterns relate to student test/exam outcomes. We extract usage information stored by the homework system (e.g. number of problem attempts and timing of the earliest correct answer) and propose metrics for student collaboration patterns in order to create “student user profiles.” We then study whether some of these profiles are associated with better grade outcomes. These results can potentially inform the implementation of WeBWorK assignments to promote those patterns.

While this study focuses on data obtained through WeBWorK, the user patterns observed would be applicable to other online homework systems, both free and paid.

2. METHODS

We collected and parsed data, derived variables from the parsed data, and performed a latent profile analysis (LPA) to identify sub-groups (also called “classes” or “profiles”) of students based on multiple variables associated with how students used the WeBWorK online homework system. LPA [7] has been used previously in post-secondary educational research [8]–[10] to find homogenous sub-groups (“classes” or “profiles”) based on a range of metrics.
2.1. Course Context

Data was collected from one integrated course with 130 students. This course covers five engineering science topics, and has a credit value equal to five regular courses. Students completed ten weekly homework assignments on WeBWorK with questions from all five subjects, as well as eleven pre-laboratory assignments and four statistics assignments. The assignments had a mix of calculated answer and multiple-choice problems (the particular distribution depended on the assignment type). Multiple-choice problems typically allowed 2 attempts, while calculated answer problems allowed unlimited attempts. Only attempts on calculated answer problems were included in this study.

2.2. Data Collection

The activity of all users (administrators, instructors and students) is recorded in two log files: the login log and the answer log. The login log contains information regarding access to the WeBWorK system by the users, including date and time, whether the login process was successful, and the user’s unique system ID. The answer log includes information about the attempts made, including the date and time, the user ID, the problem set name, question number within that problem set and whether the attempt made was correct.

2.3. Variable Selection

We considered what measures that we could obtain out of the data might be indicative of differing usage patterns among students. The specific variables we derived and used are described in Table 1.

2.4. Data Parsing

Scripts written in Python (Python.org) were used to parse through the answer and login log files and extract the data to generate the selected variables. The regular expression operations library (docs.python.org) was used to match patterns in the log files, from which useful information was extracted and stored in data structures, allowing for processing.

2.5. Variable Generation

2.5.1. Sessions and Active Sessions. Sessions were defined as any activities (attempts or logins) grouped with gaps of less than 0.95 hours (57 minutes). That is, if a student logged in and submitted an attempt 40 minutes later, followed by a 2-hour period of inactivity, the one login and one attempt would be grouped as one session.

Active sessions were defined as sessions where at least one attempt was made. (They excluded sessions with only logins).

The challenge in determining number of sessions/active sessions results from a lack of information about a gap with no activity. While such a gap could indicate that a student is no longer interacting with their homework, it could also mean that a student is working through a problem that takes a long time to solve to the point where they can enter an attempt. This issue was addressed in the work of Prat and Code [2], where they compared the timing of activities (logins and attempt submissions) with activity recorded by the learning management system (LMS) gateway that the students used to access WeBWorK. They assumed that sessions would be separated by inactivity gaps with some minimum threshold and found that an inactivity threshold of 0.95 hours led to well-correlated data with the LMS system. We adopted this inactivity threshold to define our sessions.

| Variable | Description |
|----------|-------------|
| Number of attempts (nAtt) | The total attempts made by a student over the whole course |
| Number of sessions (nSes) | The number of times a student logged in (with a gap of more than 0.95 hrs since last activity) |
| Number of active sessions (nAcS) | The number of times a student logged in and made at least one attempt (with a gap of more than 0.95 hrs since last activity) |
| Number of unique problems attempted (nQues) | The number of unique problems attempted over the whole course (maximum is the total number of problems available) |
| Attempts per session (ApS) | The average number of attempts per session. |
| Attempts per active session (SubAs) | The average number of attempts per active session. |
| Unique problems per session (QpS) | The average number of unique problems attempted in a session. |
| Collaboration (Colab) | A measure of how much students collaborated on homework. |
| Timing of first correct answer (efc) | Percentage of total time left (before homework deadline) at which the student submitted their first correct answer. |
| Timing of last correct answer (elc) | Percentage of total time left (before homework deadline) at which the student submitted their last correct answer. |

2.5.2. Collaboration. Since there is no direct indicator of collaboration between the students using the WeBWorK system, we developed an estimate of whether students were collaborating or not. Because we have access to all the
submissions of any given student, it is possible to compare the time of those submissions to the time of submissions of another student. If two individual students repeatedly had near simultaneous attempt submissions for the same question(s) over the whole course, then we believe they were likely collaborating. Because there are many more incorrect submissions than correct submissions, we chose to focus on similar timing of correct submissions only. (Note that this method would not detect collaborations where students changed who they collaborated with between problems or homework sets).

Two parameters are used to define the collaboration estimation, namely the time window for colliding correct attempts, and a colliding attempts threshold. Colliding correct attempts are defined as correct attempts to a given question of a given problem set from different students which happen within a set time window from each other. For example: assume the time window is set to 5 minutes. Student A correctly attempted question 1 of problem set 1 at 3.00 p.m. on a given day, student B correctly attempted the same question at 3.03 p.m. on the same day, and student C correctly attempted the same question at 2.53 p.m. also on the same day. In this case, only students A and B have a recorded colliding correct attempt based on the 5-minute time window.

The colliding attempts threshold is a parameter aiming to reduce the risk of false positives, which can arise when students coincidentally answer the same question correctly within the time window. This is a minimum threshold of recorded colliding attempts for two particular students to be identified as potential collaborators. For example: if the threshold is 50, then students A and B need to have 50 colliding attempts to be considered potential collaborators.

We considered time windows of 5, 10, and 15 minutes for examination. To determine the threshold number, we examined the maximum collaborators per student manually and selected the elbow point at the 5-minute time threshold (see section 3.1). We examined the average number of collaborators per student and the percentage of students who had collaborators at this point as well.

2.6. Data Analysis and Statistics

The variables in Table 1 were populated for each student and used to extract profiles using latent profile analysis (LPA) [7]. This statistical method identifies several homogeneous student sub-groups (or “classes” / “profiles”) based on a range of metrics; each sub-group has a unique pattern of metric values that can be used to develop an overall picture of that group. Students are then grouped into a number of sub-groups suggested by the LPA after setting constraints.

The sub-groups identified by LPA were compared using an ANOVA to determine if the average course-weighted tests/exams grades were different between sub-groups. (Note that the variables listed in Table 1 do not include grade information.)

All statistical analysis was performed in R, using the package “tidyLPA” [11] and baseR [12].

3. RESULTS

Our LPA analysis identified distinct student profiles based on multiple WeBWorK usage factors. We found that these profiles predicted statistically significant tests/exams grade differences.

3.1. Collaboration

We examined the number of colliding correct attempts to select a threshold for the value of colliding attempts that defines whether a student was collaborating with others or not. The correct threshold selection is important to capture true collaborations (e.g. two students working on homework together) and exclude random chance (e.g. two students who happened to work on their homework at the same time, but were not working together). We primarily considered the maximum number of collaborators a student had.

Figure 1 shows the maximum number of collaborators that any student in the class had for each potential colliding attempts threshold using a 5-minute submission time window. We found that increasing the colliding attempts threshold reduced the maximum number of collaborators to three over a wide threshold range, while a lower colliding attempts threshold increased the maximum number of collaborators to 16.

For the 5-minute window shown in Figure 1, we selected the elbow point corresponding to 4 maximum collaborators and a threshold of 30 colliding correct attempts over the course. This threshold was associated with an average of 1.1 collaborators per student, and 59% of students having collaborators. Based on instructor and student co-author experience, these values seemed reasonable.

We selected 10- and 15-minute thresholds to also provide 4-5 collaborators maximum, about 1 collaborator per student, and about half of students having collaborators (50/60 and 70/80 colliding correct attempts, respectively). These final thresholds from the three time windows (5, 10, and 15 minutes) were found to be not independent, and were combined to provide a collaboration score for each student, based on average number of collaborators, that was included in the latent profile analysis.
We varied the threshold for the number of colliding correct attempts over the whole course that would define when any two students were collaborating, and then found the average number of collaborators per student, the maximum number of collaborators any student would have, and the percentage of students with collaborators, given that threshold. Thirty colliding correct attempts at 4 maximum collaborators was the elbow point.

### 3.2. LPA and ANOVA Results

After exploring a wide range of options (with/without equality constraints on variances and covariances × number of classes from 1 to 11), we chose a three-class LPA solution with the constraints of equal variances and equal covariances mainly based on the Bayesian Information Criterion, which penalizes overfitting. We found that membership in a sub-group/class, based only on online homework variables, predicted statistically significant differences in test/exam grade in a subsequent ANOVA ($F(2, 120) = 14.59, p < 0.001$). The actual grade differences and number of students in sub-groups/classes are shown in Table 2. Post hoc comparisons using Tukey’s HSD (honestly significant difference) test suggested that the difference between Classes 1 & 3 and that between Classes 2 & 3 were both statistically significant (for both, $p < 0.001$).

### Table 2: Three-class LPA results for weighted tests/exams grades. Mean (SD). * Significantly different from both other classes ($p < 0.001$).

| Class     | Grade out of 100 |
|-----------|-----------------|
| Class 1 (n = 17) | 66.81 (10.92) |
| Class 2 (n = 81)  | 69.69 (9.99)   |
| Class 3 (n = 25)  | 80.68 (7.65)*  |
The specific profiles generate a picture of the usage patterns for each sub-group/class (Figure 2). Class 1 (lowest average tests/exams grades) has many fewer sessions and fewer overall questions attempted, but a higher number of attempts and unique questions attempted per session. Students in Class 1 have lower collaboration and later first correct answers than other classes. This paints a picture of students who might leave working on their assignment to later in the homework work period, which results in mass practice (cramming many attempts and questions into few sessions). They may have less opportunity to collaborate with others as a result of starting later (others have already done the work, or they do not have time to arrange such collaboration), and may not have had time to attempt all of the homework problems.

Class 3 (highest average tests/exams grades) shows a pattern that is fairly similar to Class 2 (middle average grade) and Class 3 (highest average grade) were substantially similar, with the largest differences occurring in the timing of first and last correct answers. While earlier work has indicated that timing of attempts was correlated with grades [3], [4], we did not expect this to remain the single major difference when other factors were included. We do not expect that this relationship between earlier correct answers and grades is causal. However, this study suggests that attempt timing may indicate a real difference in the learning practices between groups of students.

4. DISCUSSION AND CONCLUSIONS

Based on usage data obtained from our online homework system, we found three distinct profiles for sub-groups (“classes”) of students within one engineering course. These profiles were associated with statistically significant differences in tests/exams grades.

It was surprising that the profiles of Class 2 (middle average grade) and Class 3 (highest average grade) were substantially similar, with the largest differences occurring in the timing of first and last correct answers. While earlier work has indicated that timing of attempts was correlated with grades [3], [4], we did not expect this to remain the single major difference when other factors were included. We do not expect that this relationship between earlier correct answers and grades is causal. However, this study suggests that attempt timing may indicate a real difference in the learning practices between groups of students.
It was also surprising that Class 3 (with the earliest first and last correct answers) had similar numbers of attempts to the other two classes. It may be that further examination of the quality and spacing of incorrect attempts is required. For example, if a student submits an attempt, then submits two more in short order by multiplying or dividing by a factor of 10, these would not be the same quality of attempt as a student checking their work and revising their answer.

We attempted to create a measure of collaboration between students using data from the WeBWorK system. One limitation of the collaboration measurement is not having corroborating data, such as survey self-reports of collaboration. However, when spot checking results, the highest collaborating pairs were known to be friends by their instructor, so we expect that the method is picking up true collaborations to some extent. Past work found that 84% of students in three introductory physics courses using an online homework system reported collaborating with other students to some degree [4]. Our metric led to about half of students deemed to have been collaborating — this may be an underestimate which could have been caused by, say, students changing collaborators over the length of the course (which would reduce the apparent number of collaborations detected, as we examined student pairs over the whole term). Our results showing that lower average collaboration was associated with the lowest-grade subgroup is consistent with Kotas and Finck, who found that students who reported no collaboration had lower grades than those at other levels of collaboration [4]. Further exploration and validation of collaboration metrics is planned, including with student surveys.

Strengths of the methods used in this study include obtaining complete data sets of every log-in and submission completed on the online homework system for every student in one large undergraduate class. We also used an analysis method, latent profile analysis, that has been previously employed in evaluating post-secondary students over multiple factors related to learning.

Limitations include using an existing session length metric defined in another population, although that population included engineering students. We also did not control for overall sessional grade point average (GPA), however the course studied in this work comprises nearly 75% of the credits in the term for these students, and the tests/exams grades comprise 50% of the course mark (i.e. ~35% of the whole term GPA), so we would expect that the sessional GPA would be highly correlated with tests/exams grades used in this study. In future work, we will attempt to control for first year GPA, where available.

We found distinct sub-groups of homework usage patterns based on multiple factors extracted from homework system data. To our knowledge, this is the first attempt to extract profiles based on online homework data. Our results may help instructors and homework system designers to encourage student practices that are associated with better outcomes on tests and exams, through incentives (e.g. built in to the system) and/or information provided to students (as a group, or individually).

Acknowledgements

This work was completed on the traditional, ancestral, and unceded territory of the Musqueam people.

We gratefully acknowledge the UBC Scholarship of Teaching and Learning (SoTL) Seed Fund for support of this project.

We would like to thank Gianni Co for his excellent work on early data collection and analysis.

References

[1] MAA, “WeBWorK Main Page,” WeBWorK Documentation Wiki.
[2] A. Prat and W. J. Code, “WeBWorK log files as a rich source of data on student homework behaviours,” [Unpublished].
[3] G. Co et al., “Student Submission Patterns in Online Homework and Relationships to Learning Outcomes: a Pilot Study,” in Canadian Engineering Education Association Annual Meeting, 2019, pp. 1–8, Paper 36.
[4] P. M. Kotas and J. E. Finck, “Collaborative learning and other successful strategies for on-line homework,” Proc. - Int. Conf. Comput. Educ. ICCE 2002, pp. 1068–1069, 2002.
[5] M. Richards-Babb, J. Drelick, Z. Henry, and J. Robertson-Honecker, “Online Homework, Help or Hindrance? What Students Think and How They Perform,” J. Coll. Sci. Teach., vol. 40, no. 4, 2011.
[6] J. Dillion-Eggers, T. Wooten, B. Childs, and J. Coker, “Evidence On The Effectiveness Of On-Line Homework,” Coll. Teach. Methods Styles J., vol. 4, no. 5, pp. 9–16, 2008.
[7] J. Magidson, “Latent class models for clustering: A comparison with K-means,” Can. J. Mark. Res., vol. 20, pp. 37–44, 2002.
[8] H. K. Ning and K. Downing, “A latent profile analysis of university students’ self-regulated learning strategies,” Stud. High. Educ., vol. 40, no. 7, pp. 1328–1346, 2015.
[9] K. V. Keefer, J. D. A. Parker, and L. M. Wood, “Trait Emotional Intelligence and University Graduation Outcomes: Using Latent Profile Analysis to Identify Students at Risk for Degree Noncompletion,” J. Psychoeduc. Assess., vol. 30, no. 4, pp. 402–413, 2012.
[10] M. R. Pearson, A. K. Lawless, D. B. Brown, and A. J. Bravo, “Mindfulness and emotional outcomes: Identifying subgroups of college students using latent profile analysis,” Pers. Individ. Dif., vol. 76, pp. 33–38, 2015.
[11] J. Rosenberg, P. Beymer, D. Anderson, C. J. van Lissa, and J. Schmidt, “tidyLPA: An R package to easily carry out Latent Profile Analysis (LPA) using open-source or commercial software,” J. Open Source Softw., vol. 4, no. 44, p. 978, 2019.
[12] R Core Team, “R: A language and environment for statistical computing,” 2013.