Using quantitative methods for a student activity analysis in an online graduate course concerning their undergraduate education

PETRA BAGO
TOMISLAVA LAUC
DAMIR BORAS
Department of Information Sciences
Faculty of Humanities and Social Sciences University of Zagreb
University of Zagreb
Ivana Lučića 3, Zagreb
CROATIA
{pbago, tlauc, dboras}@ffzg.hr http://www.ffzg.unizg.hr/infoz/hr/

Abstract: - In this paper we analyze the use of Moodle VLE by students who attended a fully online graduate-level course. Students were split up into two groups regarding their undergraduate education. The aim of this research is to determine if there is a significant difference between the two student groups concerning both time of assignment submission and number of resources and assignments (re)accessed. We conclude that no significant difference arises regarding time of assignment submission, but this appears with respect to the access to resources and assignments. Furthermore, there is a very low correlation between the time of assignment submission and: the final course scores, the number of resources and the number of assignments (re)accessed.

Key-Words: - e-learning, online learning, Moodle, log analysis, Spearman’s rank correlation, Mann-Whitney U test

1 Introduction
Moodle [1] is a Virtual Learning Environment (VLE), also known as Course or Content Management System (CMS), Learning Management System (LMS), Learning Platform (LP) or Managed Learning Environment (MLE). More concretely, it is a free web application that educators can use to create effective online learning sites, enabling students to access course files and documents almost at any time or place.

VLEs are used to handle course-related materials and to conduct online learning activities. Student engagement in a VLE can be a source for a variety of statistical and data mining analyses. The field of academic analytics deals with applying statistical and data mining tools to academic data sets in order to reveal trends and patterns, and to construct predictive models which increase student success through evidence-based interventions.

Statistical analysis can show the most frequently accessed courses and course materials [2], and even generate reports of weekly and monthly user trends and activities [3].

Dawson et al. [4] contributed to the field of academic analytics by investigating the relationship between student online learning behavior and individual student achievement orientations. The study suggests that patterns of student participation within discussion forums can be used as an indicator of student achievement orientations.

Finnegan et al. [5] examined the behaviors of students who enrolled in different types of online courses. The result of this study implied that the most effective course design may vary depending on the content and objectives of the online courses.

In their paper, Ramos & Yudko [6] concluded that learner–content interaction is the best predictor of course outcomes. Morris et al. [7] concluded that the time spent on tasks and active online participation is important for student successful finalization.

Two basic types of content materials in a Moodle course are resources and activities. The resources are meant to be used or read without any further interaction, and they may appear in different types of files. On the contrary, the activities pursue interaction; students need to interact within activities by either submitting an online assessment, or taking part in a discussion forum, for example. Assignment submissions are time and date stamped, and this information is included along with the submitted file. Hence, information about assignment grade and due date and time can be considered.

Our previous research [8] showed that, with respect to different undergraduate education, there is
no significant difference regarding student achievement. Additionally, the obtained results showed low correlation between the number of resources (re)accessed during the course and the post-test results, as well as the number of assignments (re)accessed during the course and the post-test results. Furthermore, a significant correlation between assignments and resources (re)accessed was obtained.

In order to determine the existence of a significant difference between two student groups regarding their education, three variables were used for log analysis: time of assignment submission, number of resources and number of assignments (re)accessed. Time of assignment submission was computed in minutes from the time the teacher posted the assignment until the assignment was submitted by the learner. Number of resources or assignments (re)accessed was computed as the frequency in which the learner viewed the content of each resource or assignment.

2 Methodology

The research was performed using log analysis of student activities in an Information Sciences course at the Faculty of Humanities and Social Sciences, University of Zagreb. The course is conducted as a fully online course on a graduate level. It is elective for all students at the University and lasts for 15 weeks. The course covers advanced MS Office techniques for text and language processing. The course contains 49 resources and 15 assignments, meaning that there were 3 to 4 resources included per weekly assignment. Theoretical knowledge about basic language processing was implicitly given through practice-oriented assignments. The research involved a total sample of 54 students (34 female, i.e. 63%; 20 male, i.e. 37%) attending the course during the winter semester 2010/2011. The students were categorized in two groups considering their undergraduate education: one consisting of students from the Department of Information Sciences (“information sciences students”) at the Faculty of Humanities and Social Sciences; and the other consisting of the rest (“non-information sciences students”). The “information sciences students” are classified by gender as follows: 14 female students and 15 male students. On the other hand, the “non-information sciences students” are split into 20 female students and 5 male students (see Table 1). The “non-information sciences students” have social sciences and humanities background, except three with a background in technical sciences.

Table 1 Number of students per group

<table>
<thead>
<tr>
<th></th>
<th>female</th>
<th>male</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>non-information sciences</td>
<td>20</td>
<td>5</td>
<td>25 (46%)</td>
</tr>
<tr>
<td>students (non-inf)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>information sciences</td>
<td>14</td>
<td>15</td>
<td>29 (54%)</td>
</tr>
<tr>
<td>students (inf)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>34 (63%)</td>
<td>20 (37%)</td>
<td>54</td>
</tr>
</tbody>
</table>
• Q4: Is there a significant correlation between the time of assignment submission and the post-test results?
• Q5: Is there a significant correlation between the time of assignment submission and the number of resources (re)accessed?
• Q6: Is there a significant correlation between the time of assignment submission and the number of assignments (re)accessed?

Table 2 Statistics for student groups

<table>
<thead>
<tr>
<th></th>
<th>non-inf</th>
<th>inf</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>time of assignment submission (T)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mean</td>
<td>174,619.8</td>
<td>174,194.6</td>
</tr>
<tr>
<td>median</td>
<td>125,022</td>
<td>139,401</td>
</tr>
<tr>
<td>mode</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td>min</td>
<td>29,036</td>
<td>69,096</td>
</tr>
<tr>
<td>max</td>
<td>527,297</td>
<td>613,343</td>
</tr>
<tr>
<td>range</td>
<td>498,261</td>
<td>544,247</td>
</tr>
<tr>
<td><strong>resources (re)accessed (R)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mean</td>
<td>84.68</td>
<td>63.28</td>
</tr>
<tr>
<td>median</td>
<td>74</td>
<td>57</td>
</tr>
<tr>
<td>mode</td>
<td>/</td>
<td>55</td>
</tr>
<tr>
<td>min</td>
<td>35</td>
<td>24</td>
</tr>
<tr>
<td>max</td>
<td>210</td>
<td>139</td>
</tr>
<tr>
<td>range</td>
<td>175</td>
<td>115</td>
</tr>
<tr>
<td><strong>assignments (re)accessed (A)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mean</td>
<td>171.88</td>
<td>128.9</td>
</tr>
<tr>
<td>median</td>
<td>141</td>
<td>120</td>
</tr>
<tr>
<td>mode</td>
<td>134</td>
<td>132</td>
</tr>
<tr>
<td>min</td>
<td>77</td>
<td>65</td>
</tr>
<tr>
<td>max</td>
<td>429</td>
<td>247</td>
</tr>
<tr>
<td>range</td>
<td>352</td>
<td>182</td>
</tr>
</tbody>
</table>

*Since all the values are unique, the mode cannot be computed.*

3 Research results

3.1 Research answer 1

A two-sided nonparametric analysis is performed to determine if there exists a significant difference in the distribution for time of assignment submission between the populations “information sciences students” and “non-information sciences students”. This involves the following hypotheses:

- $H_0$: $T_{inf} = T_{non-inf}$
- $H_1$: $T_{inf} \neq T_{non-inf}$

The asymptotic significance, given by a p-value of 0.4588, leads to the conclusion that the null-hypothesis at the 5% significance level cannot be rejected. Accordingly, there is no significant difference in the mean ranks between “information sciences students” and “non-information sciences students” (see Fig.1).

3.2 Research answer 2

Based on the data given in Table 2, the following assumption is made: “non-information sciences students” access resources more than “information sciences students”. Hence, a one-sided nonparametric analysis is performed to determine if there exists a significant difference in the distribution for the number of resources (re)accessed between the populations “information sciences students” and “non-information sciences students”. This involves the following hypotheses:

- $H_0$: $R_{inf} \geq R_{non-inf}$
- $H_1$: $R_{inf} < R_{non-inf}$

The asymptotic significance, given by a p-value of 0.009138, leads to the conclusion that the null-hypothesis at the 5% significance level can be rejected in favor of the $H_1$. Accordingly, there is a significant difference in the mean ranks between “information sciences students” and “non-information sciences students” (see Fig.2).

3.3 Research answer 3

Furthermore, based on the data given in Table 2, the following assumption was made: “non-information sciences students” access assignments more than “information sciences students”. Hence, a one-sided nonparametric analysis is performed to determine the existence of a significant difference in the distribution for the number of assignments (re)accessed between the populations “information sciences students” and “non-information sciences students”. This involves the following hypotheses:

- $H_0$: $A_{inf} \geq A_{non-inf}$
- $H_1$: $A_{inf} < A_{non-inf}$

The asymptotic significance, given by a p-value of 0.009138, leads to the conclusion that the null-hypothesis at the 5% significance level can be rejected in favor of the $H_1$. Accordingly, there is a significant difference in the mean ranks between “information sciences students” and “non-information sciences students” (see Fig.3).
sciences students” and “non-information sciences students”. This involves the following hypotheses:

- **H_0**: A_{inf} \geq A_{non-inf}
- **H_1**: A_{inf} < A_{non-inf}

The asymptotic significance, given by a p-value of 0.01347, leads to the conclusion that the null-hypothesis at the 5% significance level can be rejected. Accordingly, there exists a significant difference in the mean ranks between “information sciences students” and “non-information sciences students” (see Fig.3).

3.4 Research answer 4
The Spearman’s coefficient (\( \rho = 0.1907 \)) shows that the correlation between the time of assignment submission and the post-test results is very low (see Fig.4).

3.5 Research answer 5
The Spearman’s coefficient (\( \rho = 0.0298 \)) shows that the correlation between the time of assignment submission and the number of resources (re)accessed is extremely low (see Fig.5).

3.6 Research answer 6
The Spearman’s coefficient (\( \rho = -0.136 \)) shows that the correlation between the time of assignments submission and the number of assignments (re)accessed is negative and very low (see Fig.6).

The abovementioned data show that there is no significant difference in the time of assignment submission between the groups of “information sciences students” and “non-information sciences students”. When considering the number of resources (re)accessed or assignments (re)accessed, a significant difference arises in favor of “non-information sciences students”. Furthermore, there is a very low correlation between the time of assignment submission and, either the post-test results or the number of assignments (re)accessed. Additionally, an extremely low correlation exists between the time of assignment submission and the number of resources (re)accessed. This analysis shows that the time variable is inconclusive.

4 Conclusion
In this study, we explored the use of Moodle by students who attended a fully online course. Research data reveal that, with respect to undergraduate education, no significant difference arises regarding the time of assignment submission, but this appears with respect to the access to resources and assignments. Additionally, the obtained results show a very low correlation between the number of resources (re)accessed during the course and the time of assignment submission, as well as between the number of assignments (re)accessed during the course and the time of assignment submission. Previous research showed that the frequency of participation is important for successful online learning [3], and that page hits are a very reliable means to predict student success, as well [6]. Because of the existence of a significant difference between “non-information sciences students” and “information sciences students” concerning learning-content interaction, it can be concluded that “non-information sciences students” compensated their lack of background knowledge by visiting course material more often. However, further study with a broader range of variables, such as the time spent on a task or motivational aspects, is needed to substantiate our research.

Fig.1 Box plot for time of assignment submission

Fig.2 Box plot for resources (re)accessed

Fig.3 Box plot for the number of resources (re)accessed
Fig. 3 Box plot for assignments (re)accessed

Fig. 4 Correlation between time of assignment submission and post-test results

Fig. 5 Correlation between time of assignment submission and number of resources (re)accessed

Fig. 6 Correlation between time of assignment submission and assignments (re)accessed

References:


[7] Libby V. Morris, Catherine Finnegan & Sz-Shyan Wu, Tracking student behavior, persistence, and achievement in online courses,

