p-ISSN: 2184-044X e-ISSN: 2184-1489 ISBN: 978-989-53614-5-8 © 2022
https://doi.org/10.36315/2022v2end023

EVALUATION OF CONTINUOUS STUDENT FEEDBACK ON A LARGE COMPUTER SCIENCE COURSE

Erkki Kaila, & Erno Lokkila
Department of Computing, University of Turku (Finland)

Abstract

In this article, we present a computer science course with 460 participants, where special interest was directed towards student feedback and two-way communication. The 7-week course was organized as an online course with automatically assessed exercises. Each week, the students answered a short survey about their learning experience, course difficulty level, and time spent on course activities. They were also asked to answer three open questions about what they had learned, what remains unclear and how would they improve the content. Quantitative and qualitative feedback data were analyzed for all seven weeks. In addition, we analyzed the student performance by observing the final grades obtained from the course. Based on the results, it seems that there is a positive correlation between engagement and final grades. Moreover, dynamically modifying the course on the fly based on continuous feedback seems to be a beneficial mechanism. Surprisingly, the teaching staff did not find it too laborious either. Finally, we provide experiences and suggestions for other educators for utilizing continuous feedback effectively.

Keywords: *Feedback, difficulty level, students' perceptions, programming.*

1. Introduction

Facilitating two-way communication in larger courses is difficult. Still, it is important to provide the students with a functional feedback channel to improve results, motivation, and learning experience (see e.g. Leckey et al. 2001). Based on this, we designed a feedback mechanism for a large computer science course (N=460) and utilized the feedback dynamically during the course.

In this paper, we analyze the feedback and its effects by observing

1. The changes in feedback during the course,
2. The changes in perceived workload during the course,
3. The correlation between feedback and course performance,
4. The correlation between perceived time spent and course performance,
5. The functionality of support mechanisms during the course, and
6. Students' perceptions about the feedback and support.

The paper is structured as follows: first, we present related scientific work shortly, then the course and the feedback collection are described. After this, we present the data collection and our results. Finally, limitations of the study and conclusions with some ideas for future work are presented.

2. Related work

According to e.g. Richardson (2005), student feedback is generally seen as a highly useful mechanism, but there are still a number of reasons why it is not collected. Rowley (2003) lists several factors that should be considered when a feedback mechanism is designed, including encouraging students to reflect on their learning and providing students with an opportunity to express their level of satisfaction. However, as Watson (2003) states, collecting feedback is not enough - it is also important to take action according to feedback and to inform students about the action taken.

There are several examples about utilizing feedback successfully to improve course design. Mandouit (2018) presents a study where results "supported student feedback as a valuable improvement tool, and powerful stimulus for teacher reflection" and where feedback identified areas for professional learning. Similarly, Flodén (2017) states that student feedback "has a large impact on [university teachers'] teaching and helps improve courses. Brew (2008) presents a study where students "provided constructive criticism that proved useful in the evaluation and revision process". Finally, Gaertner (2014) states that students' perceptions of teaching are valid, and may (and should) lead to changes in classroom.

There are also conflicting results. For example, Kember et al. (2002) found out, that the feedback collection did not improve the quality of teaching as perceived by students. According to the authors, a possible reason for this was that the feedback was not used effectively. Moreover, the standard questionnaire used in all 25 departments may not have been flexible enough for collecting the data. Alderman et al. (2012) also state, that while feedback is generally used and valued, questionnaires often lack validity and data is used inappropriately.

3. Course description and the feedback mechanism

The course researched in this study was a computer science course called Introduction to Programming. In addition to computer science majors, there were participants from a diverse selection of majors, including for example mathematics, physics, economy, and political sciences. The course was designed to be an introductory course with no previous knowledge of programming required. Due to covid-19 restrictions, the course was organized fully online. All exercises were automatically assessed using an online learning tool called ViLLE (Laakso et al. 2018) (known as Eduten outside Finland).

Course was divided into 7 modules (see Section 4 for details). Action research is often mentioned in relation to student feedback (see e.g. Hand et al. 2001), which is natural, as action research is dependable on continuous data. Based on similar principles, at the end of each module of the course, students were asked to answer a short survey. To keep the weekly survey as short as possible, each survey consisted of only three open questions (focusing on principles listed by e.g. Rowley 2003):

1. What did you learn this week?
2. What things remain unclear after this week?
3. How would you improve this session

In addition, the students were asked to evaluate how much time they spend on the exercises during the current week and how difficult they found the content (on a Likert scale of 1 to 7).

The results of the survey were analyzed before the next lecture. The teaching staff reacted to the feedback by 1) dynamically modifying the contents of the lectures and by 2) briefly covering the questions that were mentioned the most often in the feedback at the beginning of the lecture. The difficulty level and the time spent on doing the exercises were also discussed before the actual lecture started.

4. Data and analysis

The data was collected from the course Introduction to Programming offered at the University of Turku during the fall semester of 2021. The programming language used was Python. The course was given remotely to 460 students, of whom 358 received a passing grade.

The course structure is as follows: The first week introduced variables and the selection statement. In the second week, the students were taught repetition, i.e. for- and while loops. In the third week, students learned about functions and return values. The next two weeks focused on lists, dictionaries, and other data structures. During week six, students encountered file and error handling. The final week before the exam focused on the import-statement and some of the advanced features of Python, such as slicing and list comprehensions.

The data used in this study consists of grades given to the students and weekly feedback collected after each of the seven weeks of the course. The eighth week was the exam. The feedback questionnaire asked the following questions related to this study: How difficult was this week (Likert scale 1-7; 1 being extremely easy and 7 extremely hard), How much time do you estimate having spent on this week's exercises (categorical responses from less than 2 h, to over 20 h). We also asked students to reflect on what they learned that week and additionally, for improvement ideas and constructive criticism to further improve the course in the following years.

The grade distribution for the course is described in table 1. The grade distribution is undeniably non-normal, as proven by the Shapiro-Wilks test for normality ($W=0.7528$, $p<.0001$). The course exam was highly polarizing, as most students either received the highest grade or failed.

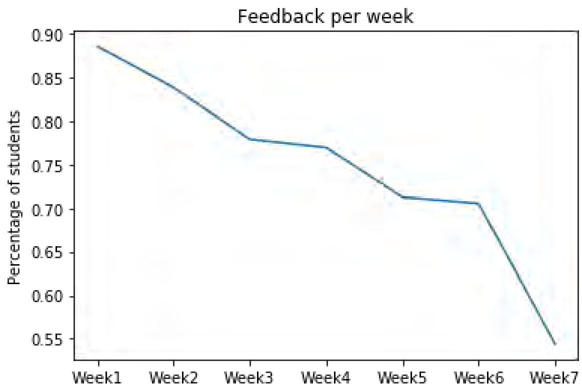
p-ISSN: 2184-044X e-ISSN: 2184-1489 ISBN: 978-989-53614-5-8 © 2022

Table 1. The grade distribution on the course.

Grade	Students	Percentage (%)
5 (Best)	215	42.74
4	66	13.12
3	38	7.55
2	25	4.97
1	14	2.78
0 (Fail)	145	28.83

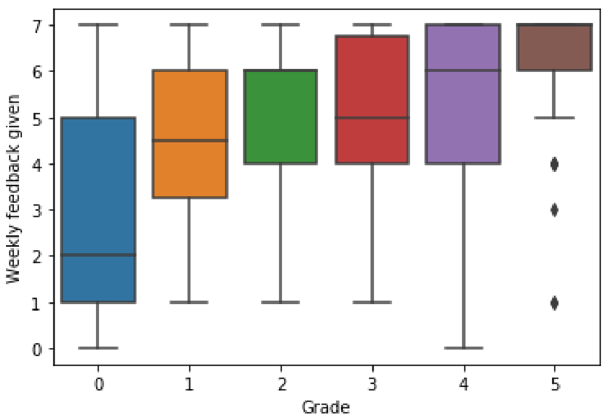
Every week students were given on average 20 small python programming assignments. The feedback questionnaire was given as the last assignment. Students received a negligible bonus on their exam scores for answering the questionnaire. As seems to be typical, the number of feedback responses dropped radically as the course went on (e.g. Barrett et al. 2018). In the first week, we received responses from 92.87% of students, but only 57.72% gave feedback the last week (Figure 1).

Figure 1. The percentage of students giving end-of-week feedback over the course.



The drop in feedback rates sparked the question: are there patterns in how students gave feedback and how did actively giving feedback reflect in the final grade. We noticed four general types of feedback givers: 1) Constant, who gave feedback every week 2) Stoppers, who actively gave feedback every week, until stopping at some point 3) Spotty, who gave feedback every now and again, with no particular pattern and 4) Disengaged, who did not give any feedback. The groups' mean grades were 4.49, 2.77, 2.71, and 0.32 respectively. We found a statistically significant difference between these groups (Kruskal-Wallis test, W=99.649, p<.0001). The difference was noticeable even without grouping the students (Figure 2). As grades were not normally distributed, we computed the correlation between these variables using Spearman's rank correlation coefficient and found a strong correlation (r=0.653, p<.0001).

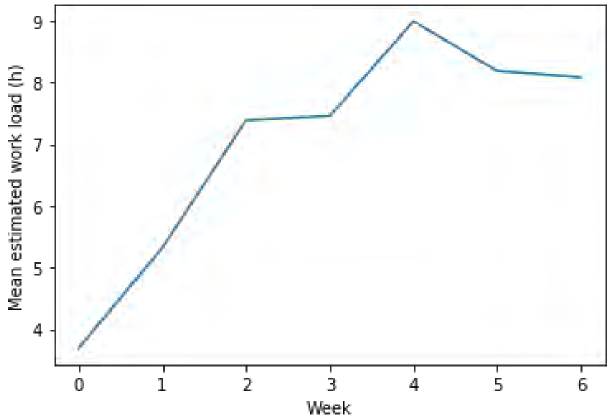
Figure 2. A Boxplot of the number of feedback for each grade.



Our questionnaire also contained a question about the perceived workload. We found a small, but significant correlation between the students' self-estimated workloads and the grades they received (Spearman's rank correlation coefficient, $r=0.343$, $p<.0001$). This means that the more students felt they worked on the course, the better their grade was.

We also had set out to find out how the perceived workload changed during the course. We found the perceived workload rose steadily, reached a high point at week 4, and then started to decrease (Figure 3). Week four focused on data structures, which either was too hard for the students or right at the point when other courses' workloads started to ramp up and students stopped putting in as much effort.

Figure 3. Students' weekly self-estimated workload in hours.



Students' perceptions about the constant feedback were mainly positive. Still, some students probably found the mechanism exhausting as seen in the decreasing number of students who answered the survey each week. In the course feedback, several students mentioned the reacting to feedback as a positive aspect:

"The lecturer listened to the students and made changes to course materials according to feedback during the course"

"I especially liked that the [student] feedback was discussed and things were revised based on the feedback".

The average grade given to the course by students (N=58) was 4.21 on a Likert scale of 1 to 5 (5 being the best), which indicates that the course was successful and students appreciated it as well. The experiences of the course staff were also positive: it seems that analyzing the feedback did not take too much time weekly, even though the course was rather large. On average, the course staff estimated that 1 to 3 hours were spent each week reading the feedback and modifying the material or preparing new examples according to feedback.

5. Limitations and future work

Naturally, there are some limitations in the study. First, we did not control for previous programming knowledge. However, it seems unlikely that all 215 students who received the best grade already had prior programming knowledge. Second, student background (major, age, gender, and so on) could have an effect on the results. Finally, we do not have a control group - it could be beneficial to compare the results to a similar course without the dynamic and constant feedback mechanism. These are all factors we are planning to include in our future studies about the subject.

6. Conclusions

In this study, we analyzed continuous feedback in a large computer science course. We found a significant positive correlation between engagement in the programming course and grades. Students who participated actively in the feedback process also gained better grades. Moreover, there was a positive correlation between students' perceived workload and their final grades. According to course feedback, the students seemed to appreciate the continuous feedback and the teaching staff's reactions to it. As it also seems that the feedback cycle does not take more than three hours per course week, we can recommend a similar mechanism to be used in other courses as well. Still, some kind of "reward" for answering the weekly survey should be utilized; gamification may also play an important role here.

References

Alderman, L., Towers, S., & Bannah, S. (2012). Student feedback systems in higher education: A focused literature review and environmental scan. *Quality in Higher education*, *18*(3), 261-280.

Barrett, M., Georgoff, P., Matusko, N., Leininger, L., Reddy, R. M., Sandhu, G., & Hughes, D. T. (2018). The effects of feedback fatigue and sex disparities in medical student feedback assessed using a minute feedback system. *Journal of Surgical Education*, *75*(5), 1245-1249.

Brew, L. S. (2008). The role of student feedback in evaluating and revising a blended learning course. *The Internet and Higher Education*, *11*(2), 98-105.

Flodén, J. (2017). The impact of student feedback on teaching in higher education. *Assessment & Evaluation in Higher Education*, *42*(7), 1054-1068.Kember, D., Leung, D. Y., & Kwan, K. (2002). Does the use of student feedback questionnaires improve the overall quality of teaching?. *Assessment & Evaluation in Higher Education*, *27*(5), 411-425.

Gaertner, H. (2014). Effects of student feedback as a method of self-evaluating the quality of teaching. *Studies in Educational Evaluation*, *42*, 91-99.

Hand, L., & Rowe, M. (2001). Evaluation of student feedback. *Accounting Education*, *10*(2), 147-160.

Laakso, M. J., Kaila, E., & Rajala, T. (2018). ViLLE – collaborative education tool: Designing and utilizing an exercise-based learning environment. *Education and Information Technologies*, *23*(4), 1655-1676.

Leckey, J., & Neill, N. (2001). Quantifying quality: the importance of student feedback. *Quality in Higher Education*, *7*(1), 19-32.

Mandouit, L. (2018). Using student feedback to improve teaching. *Educational Action Research*, *26*(5), 755-769.

Richardson, J. T. (2005). Instruments for obtaining student feedback: A review of the literature. *Assessment & evaluation in higher education*, *30*(4), 387-415.

Rowley, J. (2003). Designing student feedback questionnaires. *Quality assurance in education*.

Watson, S. (2003). Closing the feedback loop: Ensuring effective action from student feedback. *Tertiary education and management*, *9*(2), 145-157.