Exploring the Use of Peer Review in Large University Courses

Naemi Luckner, Peter Purgathofer
Vienna University of Technology (TU Wien)
Argentinierstr. 8/187, 1040 Vienna, Austria
{naemi, purg}@igw.tuwien.ac.at

Abstract. Double blind peer review is a standard practice in the scientific community. It acts as a means of validating work as well as of getting feedback to improve it. As such, it seems prudent to also use it as a learning tool in large lectures to provide students with personalized feedback on their work. The general process can be directly adopted for the lecture context, but details need to be modified and adapted to create a better learning experience. The structure of a large lecture has been adjusted to provide the context for a double blind peer review process. Not only has the evaluation of activities during the semester changed to fit in with the double blind peer review, but also the organization of said activities was adapted to accompany the evaluation change. The first semester yielded promising results, but also pointed towards some issues with the current state of the system. We devised a list of design implications for future revisions of the double blind peer review system, based on feedback and experiences during the semester as well as on a survey among students at the end of the semester. These implications will be implemented to improve and refine the new system for upcoming semesters.

Keywords: Peer assessment; Peer review; Self-directed studies;

1 Introduction

At the Vienna University of Technology, we are often facing large courses of up to 800 students. Our objective is to provide these students with personal feedback on their work, and to try to support them in increasingly self-directed studies. We want to create an opportunity for the lecture staff to take on an increasingly supportive role, rather than to pursue the more traditional power relation that results from the conventional role allocation between students and teachers. In previous semesters, we explored various approaches to achieve this goal; however, the gap opening between the large number of participants and the attempt to give detailed formative feedback regularly led to a massive backlog in evaluation of student's work.

To deal with the large amounts of students, or more specifically, student’s work, we devised a system to assess students’ work that is based on the practice of double blind peer review in scientific publication. Having students assess aspects of each other’s work lightens our workload, but still provides them with personalized and timely feedback. Additionally, students are confronted with basic concepts of scientific work and gain skills in reviewing work as well as giving and receiving productive feedback early in their studies.

In this paper, we describe the version of the double blind peer review algorithm that was used in the context of a lecture with 350 students in summer semester 2014. After experiencing the adoption of the new approach as successful during the semester, we conducted quantitative and qualitative evaluations to better understand what worked well, and what needs to be revised for future applications. The evaluation resulted in a list of design implications that we will apply to the current system for better results.

Our research goal is to explore possible spaces of organization and interaction that are opened up on the one hand by new technologies, new ideas and new paradigms, and on the other hand by new expectations, new tools and new literacies available to our students. The overall goal is not just a better or more open teaching and learning environment, but the search for new forms of teaching and learning that become possible in a situation where several hundred students push into our lectures. While we refer to Hevner et al. [1] for our general approach to design-based research, we have to interpret their ideas because we build and evaluate not only artifacts, but also amalgams of artifacts, organizational structures, and rules.

The remainder of this paper will be structured as follows: to better understand the peer review process and its application in learning contexts we first look at related work on the topic. In the following section, we present the adapted review algorithm to paint a picture of the subtle changes and additions to the generic scientific peer review algorithm. To provide context for this work, we thereafter included a chapter about *Implementation and Interface Design* of the system the peer review is integrated in. The quantitative and qualitative evaluations listed in the next section point to advantages and disadvantages of the current approach, which are addressed in detail in the *Design Implication and Discussion* Section. We conclude this work by contemplating lessons learned especially concerning our development process and the sustainability of an e-learning system in modern development environments.

2 Related Work: Double Blind Peer Review in E-Learning

Peer reviewing in various forms and configurations has successfully been implemented in university courses in the past, e.g. [2, 3, 4, 5]. The usual academic model of peer reviewing is often used in courses where students produce long texts, e.g., lab reports [6], science reports [7] or essays [8]. Other successful uses of peer reviewing outside of academia have been mirrored in education, such as code reviewing in software engineering [9].

While much related work on peer reviewing in education describes systems and processes used and designed for individual courses and settings, and offers reflection on the lessons learned, the effects of peer reviewing on learning are much harder to assess. In Pelaez work [8], the performance of students from a peer review group in a test is shown to be superior to the performance of the control group. At least one study [7] has found that students who receive peer review do a better job in revising their own work. More often, the evaluation comprises of surveying the experiences and opinions of the participating students. While some studies describe all together positive results, with students finding it to be a useful, satisfying addition to their course work [10, 11], at least one study reported the necessity for external motivation to contribute to the reviewing process [12]. Bauer et al. [13] compared modes of peer review and analyzed students' opinions towards it. Trahasch [14, 15] has been studying the effects of peer reviewing as a collaborative learning tool, finding that, on average, students did not find peer assessment particularly helpful. Lundquist et al. [16] described the positive effects of the use of a double blind peer review system in a lab course and on the skills development of students.

To enhance the effects of peer reviews, other strategies have been tried: Hart-Davidson et al. [17] devised an algorithm to measure the helpfulness of a review; Al-Smadi et al. [18] implemented a coloring schemes to easily identify problematic text parts in the reviewed work; Dominguez et al. [19] looked at strategies for successful peer-review, such as training students to give good feedback, or providing good assessment examples. Crespo et al. [20] discussed a concept where work is not randomly assigned to reviewers, but that uses an adaptive algorithm that regards user profiles and fuzzy classification techniques to reach a better assignment decision, with the goal of maximizing the benefits. Vanderhoven et al. [21] found that students had a more positive experience and felt less peer pressure in peer reviewing when the process was anonymous. Papadopoulos et al. [22], on the other hand, see an advantage in letting students freely select peers' work for feedback.

Finally, our work connects strongly with the discussion about formative vs. summative assessment. We often find that our replacement of summative assessment with traditional as well as novel forms of formative assessment best describe the design goals of our efforts. Black and William close their seminar literature review on formative assessment for learning [23] with the conclusion that 'the research reported here shows conclusively that formative assessment does improve learning'. They also point out that presently there is no optimum model to follow. A more recent review of current literature by Lau [24] challenges the conclusions of Black and William, and shows how recent research tries to re-connect these two forms of assessment so as to complement each other more than they exclude each other.

3 Double Blind Peer Review System

As a basis for the double blind peer review process, large, in-depth challenges were designed and then broken up into smaller tasks that build upon each other (Fig.1, top). We called the resulting stack of tasks a 'challenge'. Within each challenge, students have to tackle the tasks sequentially, working their way up to get to higher level tasks and finally access the final task of a challenge. Staff members grade only this last task in a challenge, thus awarding the students points for the whole challenge. By breaking up complex exercises into smaller tasks we hope to lower the initial barrier of embracing a new topic and give students more security in their proceedings.

To give an example, such a challenge could cover the technique of using interviews in usability evaluations. The first task would consist of the definition of a guide for a semi-structured interview, the second task would encompass the realization of the interview, and in the third and final task participants would do a thorough analysis of the interview, leading to an interpretation of the results. Another example could be devised around the concept of the simplification of tasks, starting with a research question (find an everyday task that is too difficult because of a badly designed product), proceed with an analysis and reflection task (describe a temporary fix or workaround that makes the task easier), and close off with a redesign task (how would you change the badly designed product so that the task becomes simpler). Each submission consists of a text field for the elaboration and can include additional material such as uploaded images and PDFs, or links to audio files and videos. After each step the peer review provides feedback by colleagues.

After finishing one task of a challenge, and before being able to move on to the next task, each student has to assess and review three of their colleagues' elaborations. All three of these elaborations originate from the same task the student has just solved, so they should be familiar with their content and requirements. The reviews are double blind, meaning the student does not know whose work they are reviewing, and the reviewee has no chance to know who has reviewed their work. Just as students are reviewing other tasks, their own tasks are being reviewed by colleagues.

In the beginning of the semester, students are given general guidelines of how to write helpful reviews. These guidelines explain how the interface and reviewing process work; how the grading system is meant to be used; how to look for plagiarism; how to write productive feedback; that reviews should be addressed to their colleagues, rather than the staff members; that good reviews take time and should not be completed as an afterthought; and that all questions of reviews have to be answered in order to hand in a complete review.

Each review is structured. Students are asked to provide formative, constructive feedback for their peers, based on a couple of specified questions related to each particular task that provide guidance. Additionally, reviewers also have to give summative feedback by categorizing the work as 'great work', 'good work', 'barely acceptable', or 'unacceptable', the latter being the only category that implies rejection (Fig.1, bottom). This 4-level assessment structure was chosen in deliberate contrast to the 5-level grading system the students are used to. Great work might earn an extra point; good work is a straight pass; barely acceptable work is missing requirements; work assessed as ’unacceptable’ will force an examination by a staff member; as long as a student has a task in her challenge that is marked as ’unacceptable’, their work in this particular stack of tasks is stopped until the situation has been cleared up. Due to the fact that they have seen three elaborations from other students, we cannot let students hand in amendments or corrections for their elaboration, or we would open the door for an easy to make, hard to detect exploit of the system.

The reviews are shown in the context of the students' task elaboration. If students do not agree with the review, the summative feedback or the grading, they can react by leaving a comment with their work. Only the lecture staff can read and react to these comments, and can help clear up issues for the students, or change the grade.

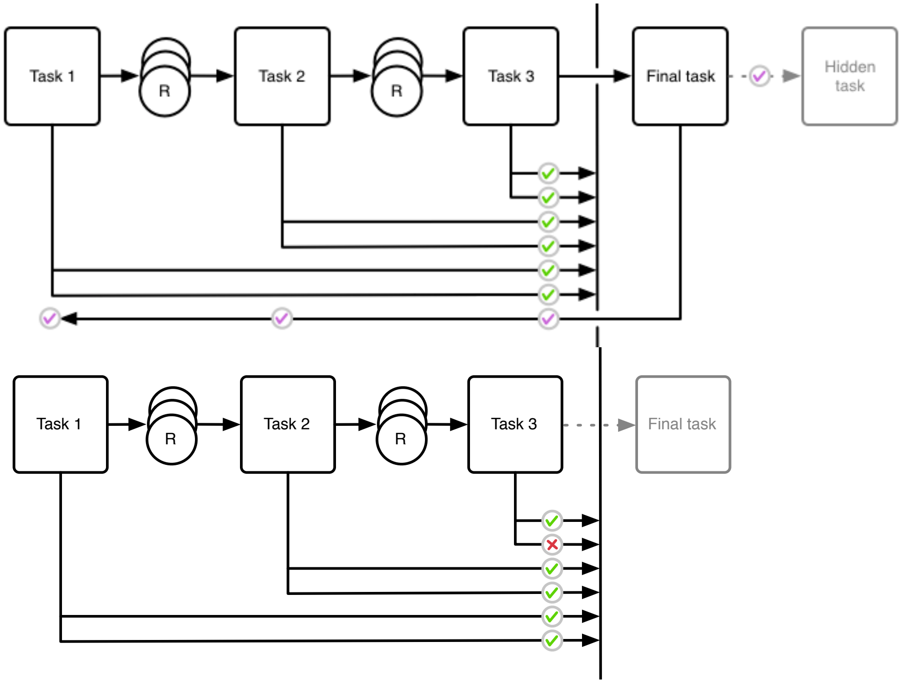


Fig.1: Multiple tasks lead up to one final task. After each task, 3 reviews (R) of colleagues' task elaborations have to be completed before moving on to the next task. Green checkmarks and red crosses are colleagues' reviews on the student's tasks, pink checkmarks are staff members' verification. To be able to hand in the final task of a challenge, a student needs to receive 2 positive reviews to each of their tasks. The top version shows a successful student, the bottom version shows a student that gets an 'unacceptable' rating in the peer review of their third task. This blocks them from being able to hand in the final task. Possible hidden tasks can be unlocked by excelling at a challenge.

Students can only access the final task once they have at least two out of three reviews on each of their previous tasks, and none is assessed as ’unacceptable’. The final task is evaluated by a staff member who has access to all of the students’ preparatory tasks in the challenge, the received reviews, as well as the reviews the student wrote for their colleagues in the context of this challenge. While the final grade is predominantly based on the quality of the final task, it takes into account the preparatory tasks, as well as of the quality of the student’s written reviews. If students repeatedly write minimalistic, meaningless reviews, they will not get full marks on their challenge, since reviewing is part of their assigned work and needs to be done properly.

When a challenge is finished and has been graded, hidden tasks can be unlocked. These tasks don’t have to be solved in order to finish a challenge, but can result in extra points upon completion. They provide students with a chance to immerse themselves deeper in a subject, and offer a chance to reward students for outstanding achievements.

Figure 1, top shows the structure for a challenge with 4 tasks and one hidden task. Green check marks represent positive reviews by fellow students, pink check marks grading by members of the course staff. A final task can be handed in after the student received 2 positive reviews to each of their tasks. The hidden task is only unlocked if all other tasks have check marks and the challenge is evaluated with full marks.

Since we still aim at enabling the students to pursuit self-directed studies, students can choose the challenges they want to work on from a large pool that offers more than double the amount of work they have to deliver. To ensure that each student gets at least an overview of all subjects inherent to the course, we created chapters that we want students to engage with. Challenges are assigned to these chapters. In a course for example, these chapters could include categories such as 'theoretical background', 'research methods' or 'evaluation methods'. Each student has to finish at least one challenge from each of the different chapters. To make sure that students get timely feedback on their work despite this open format, staff members complete missing reviews for tasks that have not been reviewed for more than 3 days. Thus, assigning the lecture staff the role as a safety net alleviates the dependency on colleagues to work on similar tasks at the same time. However, the extra workload for the staff members is manageable since there is an imbalance between the amount of written reviews required from each student to continue to the next task (three) and the amount of received reviews required per task to hand in the final task (two).

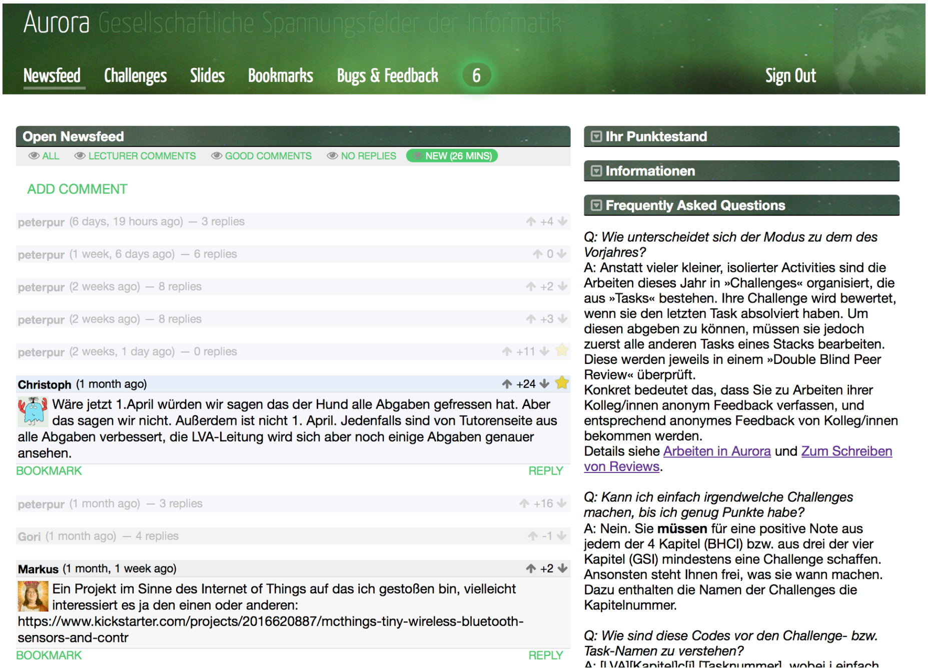
We introduced elaborate measures to counteract some anticipated forms of abuse. For example, for every task, fake elaborations showing specific weaknesses are uploaded into the system, and are randomly assigned to students. This allows for random checks of review quality. As ethical standards call for a full disclosure of this practice, students know that any task they review could be fake, which in turn serves as an incentive to maintain a good review quality.

4 Implementation and Interface Design

The system is implemented in the Django web application framework, using publicly available components such as jQuery, DropZone, or tinyMCE, to make specific functionalities and interactions possible.

The double blind peer review system is part of a larger online learning support system, *Aurora,* which was designed and implemented at our Institute (Fig.2). Other than the peer review component, *Aurora* consists of modules that facilitate announcements for and communication with the students, provides the lecture slides for reviewing and commenting, and lets students see their accumulated points and grades so far. For further information on the complete system please refer to [25, 26].

Students can find a list of all challenges (Fig.3, top left) under the corresponding top menu item. After choosing a challenge from this list, they can review all tasks the challenge consists of, and go to a task by clicking on it (Fig.3, top right). This leads to a page listing the detailed description of the task requirements, an open Q&A section for this task, and a list of all the review questions that will be used by their peers to review their work later.



Open Newsfeed

ALL · LECTURER COMMENTS · GOOD COMMENTS · NO REPLIES · NEW (26 MINS)

ADD COMMENT

peterpur (6 days, 19 hours ago) — 3 replies +4

peterpur (1 week, 6 days ago) — 6 replies 0

peterpur (2 weeks ago) — 8 replies +2

peterpur (2 weeks ago) — 8 replies +3

peterpur (2 weeks, 1 day ago) — 0 replies +11

Christoph (1 month ago) +24

Wäre jetzt 1.April würden wir sagen das der Hund alle Abgaben gefressen hat. Aber das sagen wir nicht. Außerdem ist nicht 1. April. Jedenfalls sind von Tutorenseite aus alle Abgaben verbessert, die LVA-Leitung wird sich aber noch einige Abgaben genauer ansehen.

BOOKMARK REPLY

peterpur (1 month ago) — 3 replies +16

Gori (1 month ago) — 4 replies -1

Markus (1 month, 1 week ago) +2

Ein Projekt im Sinne des Internet of Things auf das ich gestoßen bin, vielleicht interessiert es ja den einen oder anderen: https://www.kickstarter.com/projects/2016620887/mcthings-tiny-wireless-bluetooth-sensors-and-contr

BOOKMARK REPLY

Ihr Punktestand

Informationen

Frequently Asked Questions

Q: Wie unterscheidet sich der Modus zu dem des Vorjahres?
A: Anstatt vieler kleiner, isolierter Activities sind die Arbeiten dieses Jahr in »Challenges« organisiert, die aus »Tasks« bestehen. Ihre Challenge wird bewertet, wenn sie den letzten Task absolviert haben. Um diesen abgeben zu können, müssen sie jedoch zuerst alle anderen Tasks eines Stacks bearbeiten. Diese werden jeweils in einem »Double Blind Peer Review« überprüft.
Konkret bedeutet das, dass Sie zu Arbeiten ihrer Kolleg/innen anonym Feedback verfassen, und entsprechend anonymes Feedback von Kolleg/innen bekommen werden.
Details siehe Arbeiten in Aurora und Zum Schreiben von Reviews.

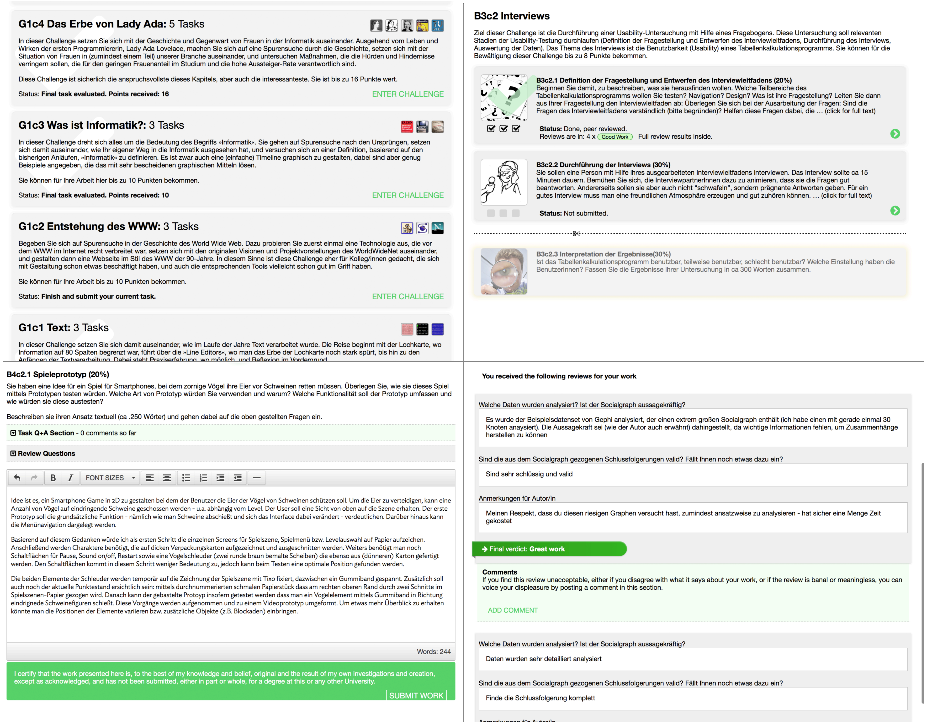
Q: Kann ich einfach irgendwelche Challenges machen, bis ich genug Punkte habe?
A: Nein. Sie **müssen** für eine positive Note aus jedem der 4 Kapitel (BHCI) bzw. aus drei der vier Kapitel (GSI) mindestens eine Challenge schaffen. Ansonsten steht Ihnen frei, was sie wann machen. Dazu enthalten die Namen der Challenges die Kapitelnummer.

Q: Wie sind diese Codes vor den Challenge- bzw. Task-Namen zu verstehen?
A: [LVA][Kapitel]c[i] [Tasknummer], wobei i einfach

Fig.2: This is what students see after logging into *Aurora*. The general layout consists of a top menu banner, featuring the main navigation on an Aurora Borealis-themed key graphics. The left column shows a newsfeed, featuring all announcements and general discussions; the right column offers static information about points and grades, the grading key, FAQ about Aurora and the current course, dates of lectures, etc.

Beneath these sections, a link offers them to ‘start this task’. Starting a task inserts the editing tools needed to hand in the work specified in the task description (Fig.3, bottom left). If only text is required, a single rich text entry field appears; if additional images are required, an extra image upload opportunity would emerge etc.

After handing in the finished work, both the list of challenges and the list of tasks change to reflect the current state of the whole challenge, using a line of text to display the status and/or explain the next step in the process. When a task has been peer reviewed, students will find the reviews they received listed beneath their work on the same page where they handed in the elaboration (Fig.3, bottom right).



G1c4 Das Erbe von Lady Ada: 5 Tasks

In dieser Challenge setzen Sie sich mit der Geschichte und Gegenwart von Frauen in der Informatik auseinander. Ausgehend vom Leben und Wirken der ersten Programmiererin, Lady Ada Lovelace, machen Sie sich auf eine Spurensuche durch die Geschichte, setzen sich mit der Situation von Frauen in (zumindest einem Teil) unserer Branche auseinander, und untersuchen Maßnahmen, die die Hürden und Hindernisse verringern sollen, die für den geringen Frauenanteil im Studium und die hohe Aussteiger-Rate verantwortlich sind.

Diese Challenge ist sicherlich die anspruchsvollste dieses Kapitels, aber auch die interessanteste. Sie ist bis zu 16 Punkte wert.

Status: **Final task evaluated. Points received: 16**

G1c3 Was ist Informatik?: 3 Tasks

In dieser Challenge dreht sich alles um die Bedeutung des Begriffs »Informatik«. Sie gehen auf Spurensuche nach den Ursprüngen, setzen sich damit auseinander, wie Ihr eigener Weg in die Informatik ausgesehen hat, und versuchen sich an einer Definition, basierend auf den bisherigen Anläufen, »Informatik« zu definieren. Es ist zwar auch eine (einfache) Timeline graphisch zu gestalten, dabei sind aber genug Beispiele angegeben, die das mit sehr bescheidenen graphischen Mitteln lösen.

Sie können für Ihre Arbeit hier bis zu 10 Punkten bekommen.

Status: **Final task evaluated. Points received: 10**

G1c2 Entstehung des WWW: 3 Tasks

Begeben Sie sich auf Spurensuche in der Geschichte des World Wide Web. Dazu probieren Sie zuerst einmal eine Technologie aus, die vor dem WWW im Internet recht verbreitet war, setzen sich mit den originalen Visionen und Projektvorstellungen des WorldWideNet auseinander, und gestalten dann eine Webseite im Stil des WWW der 90-Jahre. In diesem Sinne ist diese Challenge eher für Kolleg/innen gedacht, die sich mit Gestaltung schon etwas beschäftigt haben, und auch die entsprechenden Tools vielleicht schon gut im Griff haben.

Sie können für Ihre Arbeit bis zu 10 Punkten bekommen.

Status: **Finish and submit your current task.**

G1c1 Text: 3 Tasks

In dieser Challenge setzen Sie sich damit auseinander, wie im Laufe der Jahre Text verarbeitet wurde. Die Reise beginnt mit der Lochkarte, wo Information auf 80 Spalten begrenzt war, führt über die »Line Editors«, wo man das Erbe der Lochkarte noch stark spürt, bis hin zu den Anfängen der Textverarbeitung. Dabei steht Praxiserfahrung, wo möglich, und Reflexion im Vordergrund.

B3c2 Interviews

Ziel dieser Challenge ist die Durchführung einer Usability-Untersuchung mit Hilfe eines Fragebogens. Diese Untersuchung soll relevanten Stadien der Usability-Testung durchlaufen (Definition der Fragestellung und Entwerfen des Interviewleitfadens, Durchführung des Interviews, Auswertung der Daten). Das Thema des Interviews ist die Benutzbarkeit (Usability) eines Tabellenkalkulationsprogramms. Sie können für die Bewältigung dieser Challenge bis zu 8 Punkte bekommen.

B3c2.1 Definition der Fragestellung und Entwerfen des Interviewleitfadens (20%)
Beginnen Sie damit, zu beschreiben, was sie herausfinden wollen. Welche Teilbereiche des Tabellenkalkulationsprogramms wollen Sie testen? Navigation? Design? Was ist ihre Fragestellung? Leiten Sie dann aus Ihrer Fragestellung den Interviewleitfaden ab: Überlegen Sie sich bei der Ausarbeitung der Fragen: Sind die Fragen des Interviewleitfadens verständlich (bitte begründen)? Helfen diese Fragen dabei, die … (click for full text)

☑ ☑ ☑

Status: Done, peer reviewed.
Reviews are in: 4 x Good Work Full review results inside.

B3c2.2 Durchführung der Interviews (30%)
Sie sollen eine Person mit Hilfe ihres ausgearbeiteten Interviewleitfadens interviewen. Das Interview sollte ca 15 Minuten dauern. Bemühen Sie sich, die InterviewpartnerInnen dazu zu animieren, dass sie die Fragen gut beantworten. Andererseits sollen sie aber auch nicht "schwafeln", sondern prägnante Antworten geben. Für ein gutes Interview muss man eine freundlichen Atmosphäre erzeugen und gut zuhören können. … (click for full text)

□ □ □

Status: Not submitted.

B3c2.3 Interpretation der Ergebnisse(30%)
Ist das Tabellenkalkulationsprogramm benutzbar, teilweise benutzbar, schlecht benutzbar? Welche Einstellung haben die BenutzerInnen? Fassen Sie die Ergebnisse ihrer Untersuchung in ca 300 Worten zusammen.

B4c2.1 Spieleprototyp (20%)

Sie haben eine Idee für ein Spiel für Smartphones, bei dem zornige Vögel ihre Eier vor Schweinen retten müssen. Überlegen Sie, wie sie dieses Spiel mittels Prototypen testen würden. Welche Art von Prototyp würden Sie verwenden und warum? Welche Funktionalität soll der Prototyp umfassen und wie würden sie diese austesten?

Beschreiben sie ihren Ansatz textuell (ca .250 Wörter) und gehen dabei auf die oben gestellten Fragen ein.

Task Q+A Section - 0 comments so far

Review Questions

Idee ist es, ein Smartphone Game in 2D zu gestalten bei dem der Benutzer die Eier der Vögel von Schweinen schützen soll. Um die Eier zu verteidigen, kann eine Anzahl von Vögel auf eindringende Schweine geschossen werden - u.a. abhängig vom Level. Der User soll eine Sicht von oben auf die Szene erhalten. Der erste Prototyp soll die grundsätzliche Funktion - nämlich wie man Schweine abschießt und sich das Interface dabei verändert - verdeutlichen. Darüber hinaus kann die Menünavigation dargelegt werden.

Basierend auf diesem Gedanken würde ich als ersten Schritt die einzelnen Screens für Spielszene, Spielmenü bzw. Levelauswahl auf Papier aufzeichen. Anschließend werden Charaktere benötigt, die auf dicken Verpackungskarton aufgezeichnet und ausgeschnitten werden. Weiters benötigt man noch Schaltflächen für Pause, Sound on/off, Restart sowie eine Vogelschleuder (zwei runde braun bemalte Scheiben) die ebenso aus (dünneren) Karton gefertigt werden. Den Schaltflächen kommt in diesem Schritt weniger Bedeutung zu, jedoch kann beim Testen eine optimale Position gefunden werden.

Die beiden Elemente der Schleuder werden temporär auf die Zeichnung der Spielszene mit Tixo fixiert, dazwischen ein Gummiband gespannt. Zusätzlich soll auch noch der aktuelle Punktestand ersichtlich sein: mittels durchnummerierten schmalen Papierstück dass am rechten oberen Rand durch zwei Schnitte im Spielszenen-Papier gezogen wird. Danach kann der gebastelte Protoyp insofern getestet werden dass man ein Vogelelement mittels Gummiband in Richtung eindrignede Schweinefiguren schießt. Diese Vorgänge werden aufgenommen und zu einem Videoprototyp umgeformt. Um etwas mehr Überblick zu erhalten könnte man die Positionen der Elemente variieren bzw. zusätzliche Objekte (z.B. Blockaden) einbringen.

Words: 244

I certify that the work presented here is, to the best of my knowledge and belief, original and the result of my own investigations and creation, except as acknowledged, and has not been submitted, either in part or whole, for a degree at this or any other University.

You received the following reviews for your work

Welche Daten wurden analysiert? Ist der Socialgraph aussagekräftig?

Es wurde der Beispielsdatenset von Gephi analysiert, der einen extrem großen Socialgraph enthält (ich habe einen mit gerade einmal 30 Knoten anaysiert). Die Aussagekraft sei (wie der Autor auch erwähnt) dahingestellt, da wichtige Informationen fehlen, um Zusammenhänge herstellen zu können

Sind die aus dem Socialgraph gezogenen Schlussfolgerungen valid? Fällt Ihnen noch etwas dazu ein?

Sind sehr schlüssig und valid

Anmerkungen für Autor/in

Meinen Respekt, dass du diesen riesigen Graphen versucht hast, zumindest ansatzweise zu analysieren - hat sicher eine Menge Zeit gekostet

→ Final verdict: **Great work**

Comments
If you find this review unacceptable, either if you disagree with what it says about your work, or if the review is banal or meaningless, you can voice your displeasure by posting a comment in this section.

Welche Daten wurden analysiert? Ist der Socialgraph aussagekräftig?

Daten wurden sehr detailliert analysiert

Sind die aus dem Socialgraph gezogenen Schlussfolgerungen valid? Fällt Ihnen noch etwas dazu ein?

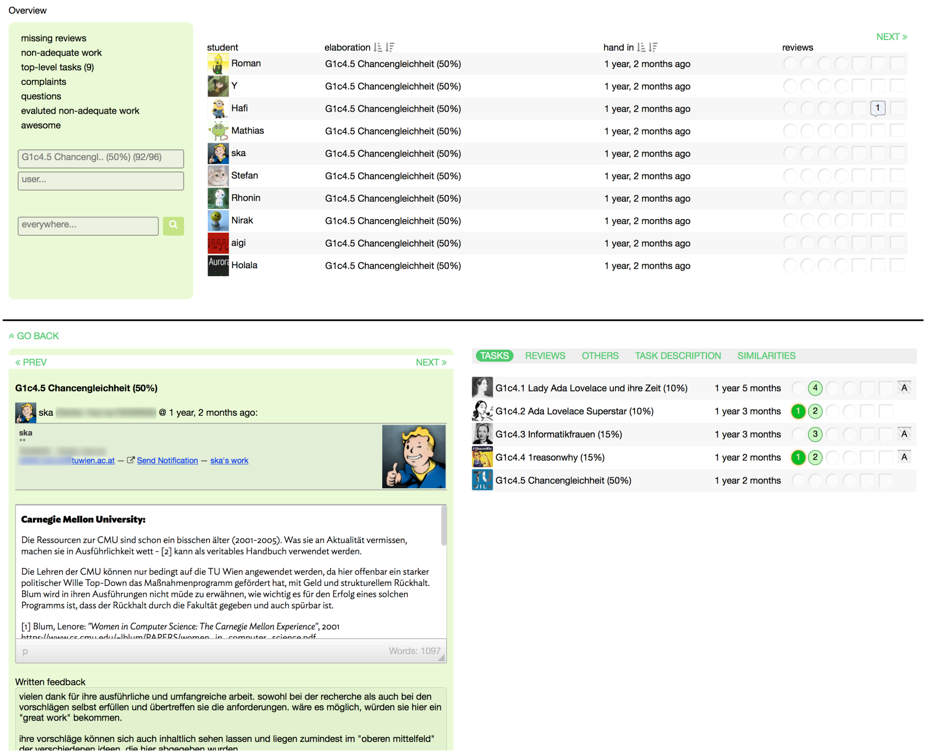
Finde die Schlussfolgerung komplett

Anmerkungen für Autor/in

Fig.3: Four screenshots from *Aurora*. Top left: list of all available challenges. Note the indication of status in the bottom left of each challenge. Top right: A single challenge under progress; the first task is finished, the second task is not submitted yet. Again, status is indicated using a text message. Bottom left: Handing in a text written for a task. Note that the Q&A-section and the section listing the review questions are both minimized to make more room for text entry. Bottom right: Display of received reviews. Note: the header was removed from all screenshots to make more room for the actual content.

4.1 Assessment View

The main menu of the assessment side of *Aurora* shows a list of recurring tasks that have to be managed throughout the semester (Fig.4, top). Items like "missing reviews" or "complaints" show all corresponding items, and contextual "NEXT" and "PREV"-links on the corresponding item pages make it easy to systematically deal with the items. Each item (Fig.4, bottom) is shown alongside a rich contextual information retrieval column, offering the possibility to display, e.g., the task description, others' students' work for the same task, or all reviews written by this student in this challenge.



student	elaboration	hand in	reviews
Roman	G1c4.5 Chancengleichheit (50%)	1 year, 2 months ago	
Y	G1c4.5 Chancengleichheit (50%)	1 year, 2 months ago	
Hafi	G1c4.5 Chancengleichheit (50%)	1 year, 2 months ago	1
Mathias	G1c4.5 Chancengleichheit (50%)	1 year, 2 months ago	
ska	G1c4.5 Chancengleichheit (50%)	1 year, 2 months ago	
Stefan	G1c4.5 Chancengleichheit (50%)	1 year, 2 months ago	
Rhonin	G1c4.5 Chancengleichheit (50%)	1 year, 2 months ago	
Nirak	G1c4.5 Chancengleichheit (50%)	1 year, 2 months ago	
aigi	G1c4.5 Chancengleichheit (50%)	1 year, 2 months ago	
Holala	G1c4.5 Chancengleichheit (50%)	1 year, 2 months ago	

G1c4.5 Chancengleichheit (50%)

ska [redacted] @ 1 year, 2 months ago:

ska
""
[redacted]
[redacted]tuwien.ac.at — Send Notification — ska's work

Carnegie Mellon University:

Die Ressourcen zur CMU sind schon ein bisschen älter (2001-2005). Was sie an Aktualität vermissen, machen sie in Ausführlichkeit wett - [2] kann als veritables Handbuch verwendet werden.

Die Lehren der CMU können nur bedingt auf die TU Wien angewendet werden, da hier offenbar ein starker politischer Wille Top-Down das Maßnahmenprogramm gefördert hat, mit Geld und strukturellem Rückhalt. Blum wird in ihren Ausführungen nicht müde zu erwähnen, wie wichtig es für den Erfolg eines solchen Programms ist, dass der Rückhalt durch die Fakultät gegeben und auch spürbar ist.

[1] Blum, Lenore: *"Women in Computer Science: The Carnegie Mellon Experience"*, 2001
https://www.cs.cmu.edu/~lblum/PAPERS/women_in_computer_science.pdf

p Words: 1097

Written feedback

vielen dank für ihre ausführliche und umfangreiche arbeit. sowohl bei der recherche als auch bei den vorschlägen selbst erfüllen und übertreffen sie die anforderungen. wäre es möglich, würden sie hier ein "great work" bekommen.

ihre vorschläge können sich auch inhaltlich sehen lassen und liegen zumindest im "oberen mittelfeld" der verschiedenen ideen, die hier abgegeben wurden

G1c4.1 Lady Ada Lovelace und ihre Zeit (10%)	1 year 5 months	4	A
G1c4.2 Ada Lovelace Superstar (10%)	1 year 3 months	1 2	
G1c4.3 Informatikfrauen (15%)	1 year 3 months	3	A
G1c4.4 1reasonwhy (15%)	1 year 2 months	1 2	A
G1c4.5 Chancengleichheit (50%)	1 year 2 months		

Fig.4: Two screenshots from the "Evaluation"-module only accessible to staff members. The left column of the top screen shows the overview-menu, offering entry points for typical tasks during the semester such as complementing missing reviews, dealing with work that was deemed "unacceptable" in peer review, or answering new questions posted in the "Q&A"-section of the task descriptions. The right column shows all work handed in for the task specified in the search field under the main menu. The bottom screen shows the item view. The left column shows a single elaboration, while the right column offers the possibility to retrieve some contextual information. Note that for space saving, the top menu bar has been removed.

5 Method and Evaluation

At the end of the semester, a voluntary survey was conducted concerning the double-blind peer review component of our assessment system. The survey consisted of 9 open-ended questions that were aimed at understanding whether the double blind peer review process was well understood, and whether the reviews provided valuable feedback to the students. Additionally, we asked for ideas for improvements to the system in future semesters.

We received results from 100 students, representing slightly less than a third of the student body. We conducted both an exhaustive quantitative and a qualitative evaluation of this data, combined with a quantitative evaluation of logged data from the system itself. On the one hand, the students' statements were thematically coded for the qualitative analysis. On the other hand, each statement was evaluated as supporting or contradicting a given question to do a quantitative analysis of the questionnaire results.

The quantitative analysis gave us insights concerning the general reception of the system as well as its usability. The qualitative analysis pointed towards concrete issues, opportunities and advantages of the system. We used those results to generate design implication for a redesign.

5.1 Quantitative Analysis

About 350 students took part in the course that is mandatory in the second semester of three bachelor programs in Informatics. In total, around 5.500 tasks were submitted into the double blind peer review process, and students wrote approximately 11.000 reviews. The staff consisted of one professor, one part-time university assistant, and 5 tutors with 5 hours per week each.

An exhaustive quantitative evaluation of data collected through the returned questionnaires, as well as collected by our system, which stores detailed information for each submitted task yielded the following information (data collected through the questionnaires is tagged as “Q”, data collected in the System is tagged as “S”):

Usability. The system was easy to understand and use for all students; some students admitted initial problems, but all of them said that after an initial period of settling in they also found the system easy to understand and use. Nobody indicated serious, lasting problems. (Q)

Learning through writing reviews. 91% of all students indicated that they learned something new about tasks they had already completed by writing a review. (Q)

Learning from receiving peer reviews. 83% of all students reported getting a helpful review at least rarely, while 17% said that they never experienced a review that was helpful for them. (Q)

'Unacceptable' work. Less than 1% (89 tasks) of all tasks submitted were rated 'unacceptable' in at least one peer review. Of those, 2/3 were assessed overly critical, in which case the course staff changed the rating. 5 tasks were blocked because of cheating attempts, 15 because of plagiarism, and 10 were unacceptable because they were really poor work, or because they failed to meet the requirements. (S)

Handling of 'Unacceptable' tasks. 75% of all students found that blocking a student from completing a challenge that has received an 'unacceptable' rating in a review is a good idea. Even from the students affected by this regulation, more than 50% agreed to this approach. (Q)

Minimalist reviewing. Low quality answers to formative review questions were a proliferating problem throughout the system over the semester. 80% of the students reported that they experienced receiving 'minimalist' reviews at least sometimes. Interestingly, students earmarked for their measly reviews in the post-course evaluation reported that they observed far less occurrences of minimalist reviews of their work. (Q) We understand this as an indication that many students practicing minimalist reviewing did not really understand the problem, and thought that their behavior was correct. As a consequence, it can be assumed that the minimalist reviewing problem can be counteracted with information and training in the right places, at the right time, as also suggested by Dominguez [19].

5.2 Qualitative Analysis

The qualitative analysis is based on the elaborate written feedback 100 students gave to 9 questions. In the following, we combined the results of this analysis with personal observations and feedback obtained through other channels during the semester to gain a better understanding of the reviewing process in the context of a university lecture. These other channels involve a lecture newsfeed with the possibility to comment, a “Bugs and feedback” section and a workshop with students who participated in the lecture. Summarized, there was a lot of positive feedback on the process, but we could also identify issues that need to be addressed in the redesign.

Remark: The quotes are taken from the survey and were in German language originally. They were translated into English by the authors.

Positive impact of writing reviews. Numerous students mentioned the positive impact that the reviewing process had on their work: "*I have to admit that I found the reviewers feedback on my work very relevant. Due to that fact, I changed from writing fairly short reviews at the beginning to rather detailed feedback, and for that reason I had more fun reading my colleagues tasks”*. Another student commented on her learning process: "*I got the most constructive feedback by writing reviews myself. Seeing how other students solved the tasks was also a kind of feedback”*. These quotes show that students did not only learn from getting reviews, but also from the process of writing reviews, engaging constructively with the course material.

Quality of reviews. For some students, reviews seemed to be more of a hassle than productive work, and they were "*annoyed by the reviews and just wanted to get them over with*" or "*wrote really short [reviews ... and] just wanted to get to the next task*". One student summarized the main issue in a nutshell: "*I only got a handful of good reviews. That demoralizes me a bit, and I caught myself getting more and more vague over time myself*". We came to the conclusion that students need incentives for writing good reviews to get them to put more work into them. Writing good reviews can act as an amplifier and motivate others to make more effort as well: "*You would need all students to write constructive feedback to make the system work”*.

Fake tasks. Fake tasks had two purposes: on the one hand, they were used as review material for the first students who hand in a given task, on the other hand they served as a quality check to find out if students take reviewing seriously. One student actually asked, "*to include more fake tasks, to single out reviewers who are not looking at the work [...] “*.

Re-framing review questions to elicit more elaborate feedback. The review questions provide students with a structure for writing a review (see Section 3). Some questions were present in every review, for example those asking for the fulfillment of all requirements, or for plagiarism. Other questions focused on the task. Some students still struggled to write meaningful feedback: "*I noticed that many had the same problem I had: some review questions could be answered with yes/no”*.

Deadlines. The system is designed to let students work at their own pace and with flexible timing as long as possible during the semester; at the same time, the system needs to account for the fact that the lecture staff cannot look at everyone's work within the last few days of the semester. We tried to put a mechanism in place that enforces a distribution of student's work over the whole semester. Students could work freely for the first half of the semester and hand in as much as they wanted, but had to pace themselves in the second half, to avoid having students hand in all their work in the last minute. Theoretically, the system would favor driven students who could be done with most of their work soon and lessen our workload at the end, but students perceived the date marking the start of the paced work rhythm in the middle of the semester as an internal deadline, and thus handed in a lot of challenges right before that date. As a consequence, the staff evaluating the final tasks was faced with a backlog that they were unable to overcome until the end of the semester.

Assessment structure. Students were generally content with the assessment categories 'great work', 'good work', 'barely acceptable', and 'unacceptable', but had some remarks. Some asked for a 5-step assessment similar to the school grading system: "*I would subdivide 'good work' into 2 categories [...] to better mark differences in quality. It might be though that I have been under the influence of the traditional grading system for too long*". Others appreciated the difference: "*An estimated assessment is reasonable because students should not evaluate work the same way the lecture staff does*" or "*[...] it is not bad having to choose between 2 [categories] and no to always be able to take the 'middle'*". There was also a discussion about the naming of the assessment categories: "*It seems unfair to assess really good work with the same category as work that does have its deficits"*.

Communicating with reviewers. Students asked repeatedly to be able to communicate with their reviewers directly: "*It is annoying that reviewers seem to not be able to read comments from authors*" or "*It would be good to consider an anonymous exchange with the reviewers to clarify misunderstandings*". The first quote refers to the possibility that authors can leave comments with their tasks even after they handed them in. Mostly these comments are used to communicate with the lecture staff, but sometimes, students add some information to their tasks, e.g., appending forgotten references. These delayed edits can be seen by the staff to include them in the final evaluation of the stack, but not by the reviewers who then comment on the missed requirements. The second quote addresses the lack of communication between reviewer and author. This was a conscious design decision that will be discussed more in-depth in the design implications section.

Review timing. Some students mentioned issues with the time they got their reviews and its impact on postponing handing in their final task: "*I think it [the review system] is great. The only problem I have with it are the long waiting times for reviews, especially in older challenges*". Since students have to wait to get the right amount of positive reviews before being able to hand in their final task, they depend on their colleagues to be on top of their work as well, or they need to wait until the lecture staff creates substitutes for said reviews, which only happens 72 hours after handing in a task. Students suggested, "*to shorten the time until the lecture team reviews tasks*", that "*one should be able to hand in the finals even without reviews*" and that "*the system could unlock the final task automatically after 72 hour*". They also offered to *"voluntarily register to write more [than 3] reviews, maybe getting extra points for 10 additional reviews"*.

Not one at a time. A few students remarked that *"It would be good to see all three tasks I need to review, to have a basis of comparison"*. The issue mentioned here refers to the fact that when students are at the review stage, and are assigned a task to review, they have to finish reviewing this task before getting their second and then third task to review. Hence, they only see one task elaboration at a time and especially for the first task only have their own work to compare it to.

Plagiarism. For every task, one of the questions reviewers have to answer is whether the work is in any way plagiarized. In the survey, students reacted to this as follows: *"Plagiarism checks should not be passed on to reviewers, since it is a pesky activity"*. On the other hand, reviewers exposed a number of plagiarized works, which made the staffs evaluation tasks a lot easier.

Cheating. We also uncovered two other ways students used to cheat the system:

One group of students choose a time, preferably when only few other students would work on a challenge (e.g. right after the requirements of a challenge were published and in the middle of the night), and would all hand in at the same time. Since the pool of new tasks was comparably small, the chance to get their friends' work to review was comparably high. They gave each other great reviews on mediocre or bad work and moved on to the next task. We were able to uncover those networks in analysis of the data afterwards, but found it hard to justify swift measures against this circumvention of our system.

The other way of cheating we discovered was a group of students who used a colleague who had already dropped out of the lecture for that semester as informant. The dropped out colleague still had access to the system, handed in close to empty tasks, started reviewing, copied the work he had to review and sent it to his colleagues. These colleagues chose from these looted tasks, sometimes made small changes and handed it in as their own. Fortunately, we were able to catch most of these incidents with a hastily implemented similarity checker.

6 Design Implications and Discussion

We regard the system as a work in progress, and try to learn from previous mistakes. The evaluation of the survey as well as our own experiences during the semester led us towards some redesign inspirations that will impact the next version of our implementation of a double blind peer review system. The following is not an exhaustive list of upcoming changes, but provides insights into the ongoing design process. Also, it shows how the evaluation of previous semesters led to design implications and influenced design decisions for upcoming versions of the system.

6.1 Design implications for the summative review

The survey pointed towards the fact that a lot of students experienced problems with the four-level summative review system, both while reviewing and when receiving reviews. This led to a major re-design of the feature with the hope to make it more accessible to the students:

Re-naming of categories. Students did not want to get the feeling that they are grading their peers, so we decided to rename the categories for the summative feedback. The new, more descriptive assessments are 'Better than my own work', 'Acceptable', 'Requirements missed' and 'Plagiarism or cheated'.

The 'Better than my own work' assessment category. Reviewers often struggled with the frame of reference that would let them define a task as 'great work': *"Especially with 'Great Work', it is hard to estimate if it can really be counted as extraordinary and in all tasks I have reviewed so far, only two can really be categorized as 'Great Work'"*. The new wording 'Better than my own work' will introduce a clear frame of reference. Students will not be required to judge if a task could be seen as 'great work' in the context of the whole lecture, but will rather compare the work to their own. Assessing a task as such will need an extra explanation that explains the differences between the reviewer's and author's work. Since reviewers will have extra work when flagging other work as ‘better than my work’, they might need an incentive to do so. As writing such an explanation is essentially a reflection on the reviewers own work and hence contributes to the reviewer's own learning process, this additional work as it’s own value that should be acknowledged and rewarded; reviewers could e.g. get good 'review karma' (see below), which will translate into bonus points at the end of the course.

The 'Acceptable' assessment category. The acceptable category encompasses all work that, in the eyes of the reviewers, fulfills the task requirements. Issues will still be addressed by the reviewers in their formative feedback to the authors and can as such be seen by the staff members, who can include it in the final evaluation.

The 'Requirements missed' and 'Plagiarism or cheated' assessment category. Students usually gave 'unacceptable' assessments if the author plagiarized, cheated or did not meet the task's requirements. These categories are grounded in the data (see the quantitative evaluation section) and were also used as justification when staff members deemed a student’s work unacceptable. However, only if work is classified as 'plagiarism' or 'cheated', the author will be blocked from further work in the affected challenge until the issue is cleared; a classification as 'requirements missed' will have no immediate consequence for the author, but might have significant influence on the overall assessment of the challenge by a staff member. Hence, we created the 'requirements missed' assessment category to show that something is missing in the task, but that the work is nevertheless acceptable.

Displaying assessment. The reviewers' summative feedback is highly valued by the lecture staff, as it helps to reduce the grading workload. However, displaying this personal assessment to the evaluated student was often not received well. Again and again, students commented about their unease with being 'marked' by their peers. They compared the assessments to the school grading system, and repeatedly deposited in comments that they should be getting better grades for their work, even though the real grading was really only done by members of the staff, considering the quality of work for the whole challenge. After carefully deliberating the issue, we decided to still ask reviewers for a personal summative assessment of the work, but to use it for internal purposes only. Students will not see this assessment along with the other parts of the review anymore. Instead, they will only see their work being 'accepted' or 'not accepted' by the reviewers.

6.2 Design implications for the reviews

The reviews themselves will need to undergo some changes, most of which pursue the common goal of increasing the quality of the reviews and avoiding bad reviews in regard to the content.

No replies to reviews. Most replies to reviews concerned the reviewers' assessments. For example, some students complained about just getting a "barely acceptable" even though they thought they did a good job. Only in a few cases did replies actually yield any constructive discussion about the review. Additionally, a lot of students thought that reviewers would be able to see replies, and thus addressed those comments directly to them, of course without receiving any satisfying answer thereafter. Since we see reviews as a means of feedback for the students on their work, and feedback does not need to be answered, we concluded that replies are not necessary. For major issues students want to communicate to the staff team, they henceforth still have the possibility to comment on their work in general, but not on individual reviews.

Promoting good reviews. We are considering different strategies to promote a practice of writing detailed, considerate reviews, such as a community-based rating mechanism of reviewers that we plan to call 'reviewer karma'. Each student rates the reviews they received as 'helpful', 'average', or 'meaningless or offensive' review. The 'reviewer karma' changes accordingly. The level will not rank the reviewers in relation to other students, but reflect on their average review practices. Following a student's suggestion, there will be positive and negative consequences for extremely good or inconsiderate reviewers, such as bonus point or deducting points from their final score: "*Good reviews should be positively rewarded, even if it's only a symbolic gesture*". If such a rating is established, it could be interesting for the reviewee to see the 'reviewer karma' of each reviewer to better contextualize the feedback received.

Dealing with bad reviews. Apart from promoting good reviews, there also needs to be a better system to deal with low quality reviews. Apart from a feedback mechanism for reviews as described above, students also suggested to "*put in more fake work (written by the lecture staff) to make students more attentive«,* with the goal to "*... block the person from continuing*" or to "*flag bad reviewers*".

Showing all tasks that need to be reviewed at once. Even tough students asked to see all tasks they need to review at the same time to be able to compare the work, we decided against it for practical reasons: As soon as a tasks is assigned to a student for a review, they have to complete the review for that exact task. This rule insures that the reviewer is chosen randomly and cannot change their assigned task for a "better" one, e.g., that of a friend. Hence, if the reviewer receives all the tasks they need to feedback at once, they have to complete all of them, so the authors of the tasks have to wait for the reviewers to do their work. In case they do not finish all of their reviews in one go, all three works they have to review will still be blocked from being assigned to another reviewer. This potentially prolongs the time students have to wait for reviews.

6.3 Design implications for the review questions

The wording and semantics of review questions need to be revised with a focus on formulating those questions to elicit more elaborate feedback from the reviewers. Students asked for "*more specific questions to make sure the reviewer really dealt with the work*" and "*open questions based on the topic to make sure the task satisfies the requirements*".

6.4 Design implications to improve work based on feedback

Understandably, students asked for a way to act on the feedback they received, and to have the possibility to improve their work based on it: "*It would be good to put a mechanism in place to revise old work, since most will just ignore the reviews as long as they are 'positive'*". Incidentally, we had already experimented with mechanisms for students to hand in updated versions of their work, but found that it generated too much extra work for the lecture staff, especially giving the high number of participants. However, providing students with the opportunity to improve their work based on the feedback is still an important feature for us, which will somehow find its way back into the system.

7 Conclusion and Future Work

For some years now we have been exploring different forms of online hand-in, evaluation and feedback to accommodate learning in a self-reliant, self-controlled and self-directed way [25, 26]. With the introduction of a massive online double blind peer reviewing process, we think we have finally found a practical and sustainable way of organizing a high-quality learning experience for a large number of participating students without compromising our goals.

The good reception and easy handling of the system by the students, as well as the comparatively low organizational overhead generated for the course administrators and teachers gave us the impression that this approach is sustainable enough to build and improve upon. The redesign suggestions discussed above affect the immediate future of the system.

More problems loom, as we look further into the future. Experimental systems like the one presented here are often developed within the context of a research project, usually with the help of students who contribute to the project in the course of their bachelor or master thesis. Consequently, such systems can suffer from 'software rot', a popular term describing how code that is left unchanged in an ever-changing environment of operating system updates, security fixes and enhancements becomes faulty to the point of obsolescence. This is certainly true for the project discussed here, and we already experience problems in older components of the overall system.

Today, the speed and intensity of software maintenance is determined by the industries' general pace of development, which some argue is quite neck braking for smaller players. We believe this to be a common problem for projects where technologies are designed, developed and deployed that could be hugely beneficial to minorities or niche markets, but cannot easily be turned into or sustained as a commercial endeavor. Nevertheless, we won't stop looking into ways to secure the long-term survival and development of our system.

References

1. Hevner, A. R., & Chatterjee, S. (2010). Design Research in Information Systems. Media. New York Dordrecht Heidelberg London: Springer.
2. Gehringer E.F.: Strategies and mechanisms for electronic peer review Frontiers in Education Conference, 2000. FIE 2000. 30th Annual. vol. 1. pp. F1B/2–F1B/7 vol.1 (2000)
3. Liu E.Z.-F., Lin S.S.J., Chiu C.-H., Yuan S.-M.: Web-based peer review: the learner as both adapter and reviewer Educ. IEEE Trans., 44, pp. 246–251 (2001)
4. Wolfe W.J.: Online Student Peer Reviews Proceedings of the 5th Conference on Information Technology Education. pp. 33–37. ACM, New York, NY, USA (2004)
5. Nagel L., Kotzé T.G.: Internet and Higher Education Supersizing e-learning: What a CoI survey reveals about teaching presence in a large online class Internet High. Educ., 13, pp. 45–51 (2010)
6. Berry D.E., Fawkes K.L.: Constructing the components of a lab report using peer review J. Chem. Educ., 87, pp. 57–61 (2010)
7. Trautmann N.M.: Interactive learning through web-mediated peer review of student science reports Educ. Technol. Res. Dev., 57, pp. 685–704 (2009)
8. Pelaez N.J.: Problem-based writing with peer review improves academic performance in physiology. Adv. Physiol. Educ., 26, pp. 174–184 (2002)
9. Garousi V.: Applying peer reviews in software engineering education: An experiment and lessons learned IEEE Trans. Educ., 53, pp. 182–193 (2010)

10 Basnet B., Brodie L., Worden J.: Peer assessment of assignment Frontiers in Education Conference (FIE), 2010 IEEE. pp. T1G–1–T1G–2 (2010)

11. Settle A., Wilcox C., Settle C.: Engaging Game Design Students Using Peer Evaluation Proceedings of the 2011 Conference on Information Technology Education. pp. 73–78. ACM, New York, NY, USA (2011)
12. Turner S., Pérez-Quiñones M.A., Edwards S., Chase J.: Student Attitudes and Motivation for Peer Review in CS2 Proceedings of the 42Nd ACM Technical Symposium on Computer Science Education. pp. 347–352. ACM, New York, NY, USA (2011)
13. Bauer C., Figl K., Derntl M., Beran P.P., Kabicher S.: The Student View on Online Peer Reviews SIGCSE Bull., 41, pp. 26–30 (2009)
14. Trahasch S.: From peer assessment towards collaborative learning Frontiers in Education, 2004. FIE 2004. 34th Annual. pp. F3F–16–20 Vol. 2 (2004)
15. Trahasch S.: Towards a flexible peer assessment system Information Technology Based Higher Education and Training, 2004. ITHET 2004. pp. 516–520 (2004)
16. Lundquist C., Skoglund M.A., Granström K., Glad T.: Insights from implementing a system for peer review IEEE Trans. Educ., 56, pp. 261–267 (2013)
17. Hart-Davidson W., McLeod M., Klerkx C., Wojcik M.: A Method for Measuring Helpfulness in Online Peer Review Proceedings of the 28th ACM International Conference on Design of Communication. pp. 115–121. ACM, New York, NY, USA (2010)
18. Al-Smadi M., Gütl C., Kappe F.: Towards an Enhanced Approach for Peer-Assessment Activities Advanced Learning Technologies (ICALT), 2010 IEEE 10th International Conference on. pp. 637–641 (2010)
19. Dominguez C., Nascimento M.M., Payan-Carreira R., Cruz G., Silva H., Lopes J., Morais M. da F.A., Morais E.: Adding value to the learning process by online peer review activities:

towards the elaboration of a methodology to promote critical thinking in future engineers Eur. J. Eng. Educ., (2014)

20. Crespo García R.M., Pardo A., Delgado Kloos C.: An adaptive strategy for peer review Frontiers in Education Conference. pp. F3F–7–13 Vol. 2. Ieee (2004)
21. Vanderhoven E., Raesa A., Montrieuxa H., Rotsaerta T., Schellensa T.: What if pupils can assess their peers anonymously? A quasi-experimental study. Comput. Educ., 81, pp. 123–132 (2015)
22. Papadopoulos P.M., Lagkas T.D., Demetriadis S.N.: How to improve the peer review method: Free-selection vs assigned-pair protocol evaluated in a computer networking course Comput. Educ., 59, pp. 182–195 (2012)
23. Black, P., & Wiliam, D. (1998) Assessment and Classroom Learning. Assessments in Education: Principles, Policy and Practive, Volume 5, Issue 1, 1998
24. Lau, A. M. S. (2014). 'Formative good, summative bad?' – A review of the dichotomy in assessment literature. Journal of Further and Higher Education, (2015).
25. Purgathofer P., Luckner N.: Layout Considered Harmful : On the Influence of Information Architecture on Dialogue in Zaphiris, P. and Ioannou, A. (eds.) Learning and Collaboration Technologies. Designing and Developing Novel Learning Experiences, HCII 2013. pp. 216–225. Springer International Publishing, Heraklion, Crete (2014)
26. Luckner N., Purgathofer P.: Explorative Design as an Approach to Understanding Social Online Learning Tools Int. J. Adv. Intell. Syst., 7, pp. 493 – 506 (2014)