Graded Team Assignments in MOOCs - Effects of Team Composition and Further Factors on Team Dropout Rates and Performance

Thomas Staubitz, Christoph Meinel
Hasso Plattner Institute, University of Potsdam, Potsdam, Germany
{firstname.lastname}@hpi.de

ABSTRACT

The ability to work in teams is an important skill in today's work environments. In MOOCs, however, team work, team tasks, and graded team-based assignments play only a marginal role. To close this gap, we have been exploring ways to integrate graded team-based assignments in MOOCs. Some goals of our work are to determine simple criteria to match teams in a volatile environment and to enable a frictionless online collaboration for the participants within our MOOC platform. The high dropout rates in MOOCs pose particular challenges for team work in this context. By now, we have conducted 15 MOOCs containing graded team-based assignments in a variety of topics. The paper at hand presents a study that aims to establish a solid understanding of the participants in the team tasks. Furthermore, we attempt to determine which team compositions are particularly successful. Finally, we examine how several modifications to our platform's collaborative toolset have affected the dropout rates and performance of the teams.

ACM Classification Keywords

K.3.1. Computer Uses in Education: Collaborative learning;
K.3.1. Computer Uses in Education: Distance learning

Author Keywords

Teamwork; MOOCs; Team-based Learning; Team Assessment; Peer Assessment; Project-based learning

INTRODUCTION

The ability to work in teams is an important skill in today's work environments. Collaborative learning, teamwork, project-based learning, active learning are becoming more and more popular as teachers and industry alike see this as a way to tackle the challenges of the future. "Learning" is becoming less and less the ability of learning facts or techniques and more and more the ability of learning how to learn and to adjust to ever new requirements. The social constructivist learning theory considers learning as a social activity. It happens whenever we interact with other humans and works best when we work on a common task. Collaborative learning among the course participants has been an important element of our platform since we started our first MOOC in 2012. Types of collaboration range from low-profile, large-scale collaboration, such as discussions about certain aspects of a quiz or a video in the general course forum to the provision of *learning rooms* for loosely coupled groups of participants. Originally, these *learning rooms* mainly provided a private discussion forum and the possibility to share files for the group members. Each participant is able to create such *learning rooms* and can decide whether the room will be public—open for everybody to join—or private—only invited learners can join. Later, for reasons that are beyond the scope of this paper, we renamed these *learning rooms* to *collab spaces* and added further communication and collaboration channels. Finally, in 2016 we conducted the first course containing a graded team-based assignment on our platform. Since then, we have conducted 15 courses in total, containing graded team-based assignments.

For our study, we have analyzed the data of eight of these courses. All of them have been completely open to the public and their size ranged from about 2,000 to 20,000 enrolled participants. The course topics can be grouped into the following categories:

1. Object-Oriented Programming in Java (OOP):
   Two almost identical iterations of the same course
   javaeinstieg2017 (German language, 252 teams)
   java1 (English language, 119 teams)

2. Business Innovation (BI):
   Enabling Entrepreneurs to Shape a Better World
   (*sbw1*, 39 teams)
   Business Models for the Digital Economy
   (*bmi1-1*, 49 teams)
   Intrapreneurship (*bizmooc2018*, 28 teams),

3. Design Thinking[1] (DT):
   Developing Software using Design Thinking

---

[1]Design thinking is a user-centered approach for problem-solving and idea development. Stanford University initially extended and developed Design Thinking education programs. The approach has been implemented in organizations internationally [3] [8].

(Three iterations)
(*dt1*, 62 teams; *dt1-1*, 66 teams; *dt1-2*, 48 teams)

For each of the courses, the instructors have selected a **suitable** task for the team assignment. In our definition, a suitable task for a team-based assignment is a task that would also be solved by teams in a real work environment.

The task in the OOP courses was to model a given problem and visualize it as a UML[2] class diagram. The deliverables have been the diagram, a short documentation of the (planned) program including a glossary of terms, and a code scaffold that translated the diagram to Java code. The requirement for the code scaffold was that it had to compile, the methods, however, did not have to be implemented. The task had a duration of two weeks. It allowed the participants to collect an amount of bonus points that corresponded to the maximum possible points for a mandatory weekly assignment.

In the BI courses, the participants had to develop and pitch a business model in the context of either digital transformation, intrapreneurship, or social entrepreneurship. The deliverables have been a slide set and an optional video. In *sbw1* and *bmi1-1*, the participants worked on the task for about six weeks and had to deliver intermediate results on a weekly basis. The points from the task formed an essential part of the overall course grade. In *bizmooc2018*, some initial tasks, e.g. the selection of the topics to be tackled by the teams, have been elaborated collaboratively with the entire course population. The actual team task had a duration of ~two weeks and was the requirement to complete the course's full track—a fast track, without a team task, was also available.

In the DT courses, a prototype e.g. for a vending machine had to be developed. During the project the participants had to apply several steps of the design thinking process. The deliverable was a slide set that documented the team's process. The duration of the team task was six weeks. Intermediate results had to be submitted on a weekly basis. The points from the task formed an essential part of the overall course grade.

In each of the courses, the final deliverables have been peer assessed by the members of other teams. In all courses, the participants had an additional week for reviews and assessment.

In a peer-assessed team assignment one of the team members submits the team's solution. Then, all team members have to review the work of other teams individually. Additionally, the team members can rate the contribution of their own team mates. Finally, the teams jointly rate the reviews they've received for their submission. Participants only receive points if they have reviewed the work of other teams. Therefore, lurkers or dropouts within the teams will not receive points for work they haven't done. To participate in the team task, course participants have to register separately and provide a set of additional data, which is used to match the teams.

With our study, we attempt to answer the following questions:

- Which are the differences (if any) between the total course population and the subset of participants that register for the team task?

- Which constellations in the composition of teams have particularly positive or negative effects on the teams' performance or dropout rates?

- How have our platform modifications affected the teams' performance or dropout rates?

The remainder of the paper is structured as follows: we start with a selection of related work in Section 2. In Section 3, the general context of our research is outlined and the paper at hand is positioned in this context. In Section 4 we examine the differences between the total course population and the subset of team members. In Section 5 we analyze the effects of team composition and other factors on the teams' performance. Finally, in Section 6 we highlight our next steps and conclude our work.

RELATED WORK
First, we present research, ideas, and solutions that have inspired our work. We have focused on the topics *team-based assignments in MOOCS* and *team building* (selecting a sufficiently good combination of team members). Although *team forming* (transforming a group of individuals into a team) and *team grading* are important aspects of our overall research, we have omitted these topics here as they are only marginally important for the research presented in the paper at hand.

Team-based assignments in MOOCs
NovoEd is one of the few major MOOC platforms that supports the concepts of teamwork and collaboration with powerful tools. Already back in 2013, they offered a MOOC with explicit team-based assignments in Spanish language [1]. Rosé et al. [11], report about a MOOC that they conducted on the edX platform, which also contained a collaborative reflection activity. In the beginning of 2016, several California community colleges announced to bring teamwork to their online classes. Bazaar[3], a tool to support discussions in teams by introducing an AI agent, which triggers and guides conversations among students was to be employed for this purpose [10]. In Berkeley's *Engineering Software as a Service* MOOC in 2014, the participants were asked to use Google Hangouts to work on ad-hoc pair-programming sessions [9]. Ju et al. [6] are developing a tool to support MOOC instructors in coaching teams in agile software engineering courses.

Team-building - matching team members
We adapt Kizilcec's [7] wording for team-building approaches: *interventionist* for teams that are built by the instructors and *laissez-faire* for teams that are built by the participants' themselves. Interventionist instructors might form teams either randomly or based on a selection of well-defined criteria, which can be applied either in a homogeneous or a heterogeneous way. Whether the instructors decide on a random or a criterion-based approach, depends on the number of learners that have to be "teamed" and the tools that the

---

[2]UML: Unified Modeling Language

[3]http://www.cs.cmu.edu/~cprose/Bazaar.html

instructors have at hand to support them forming the teams. If no such tool is available, the laissez-faire approach might appear to be the easiest solution for the instructors. However, there are some arguments against it:

- Teams consisting of friends seem to perform better on tasks with high quantity output. They perform worse, however, in tasks with high quality output. Teams of strangers are stronger in "constructive disagreement" [4].

- Teams that consist of some members who already know each other and others who don't, are more vulnerable towards the formation of subgroups within the team. Subgroups have a negative influence on the performance of a team and can frustrate team members who are not part of the subgroups [5].

- The team building process itself might be frustrating or humiliating for some participants when they are rejected or leftover.

- The team building process requires a substantial amount of self-confidence and extroversion as the participants have to reach out to others.

- Particularly in MOOCs, with thousands of learners, who do not know each other at all, the laissez-faire approach is hard to manage for the participants.

Kizilcec [7] confirms the need for other ways to build teams than random selection or a laissez-faire approach. Shimazoe and Aldrich also discourage the laissez-faire approach in the team building process [12].

Wen [16] proposes the following approach to form teams out of a crowd of strangers: First, participants submit an individually produced artifact in thread of the course wide discussion forum. There, the submitted artifacts are discussed by the course participants. Finally, the forum interaction is analyzed and students, who have engaged in meaningful discussions are teamed up. Wen's experimental results indicate that the groups that have been formed with this approach produce a better learning outcome than those that are randomly formed [16]. Zheng and Pinkwart [17] propose a matching algorithm that allows to dynamically re-compose teams, based on the teams' performance and the students' satisfaction. After several iterations, the best-performing teams have been matched. Belbin [2] identified nine archetypical team roles and developed a test to determine for which of them a team member might be best suited. Teams are composed with the goal to have all roles represented in each team.

In our particular context, building the teams has to be very quick and efficient as we only have a very limited amount of time for this task. The total length of our courses doesn't exceed six weeks, the team tasks are often part of even shorter hands-on courses or workshops with a length of two to four weeks. We have a maximum timeframe of one or two days between the deadline of the registration and the start of the team task to build the teams. The teams work on the same project for the whole duration of the task. Re-composing the teams once the task has started, is not an option in most cases. Therefore, we can neither rely on lengthy processes as the one described by Wen, nor on multiple iterations as

proposed by Zheng and Pinkwart. Furthermore, the bond between our platform and our learners is by far not as close as the one between a regular university student and her alma mater (or between employee and employer). Hence, extended questionnaires as proposed by Belbin are also not suitable for our purposes.

RESEARCH METHODOLOGY

The current work is a part of a long-term study, which follows an iterative mixed methods approach. Having gained experience in peer assessment and the support for collaborative learning on our platform, we combined the two, and, in 2016, delivered the first MOOC on our platform that contained a graded team-based assignment (*sbw1*).

Technically these assignments rely on three elements:

- The platform's *collab spaces*. They provide a selection of communication and collaboration tools and originally have been developed to enable the participants to form self-organized learning groups. The *collab spaces* for team tasks have been modified so that every team member is always informed per email when new posts are added in the team forum.

- The platform's peer assessment system. It has been extended to (1) allow submissions not only by individuals but also by teams; (2) allow individual team members to review and grade the work of other teams; (3) allow the team members to rate the contribution of their team mates.

- An additional tool—the Team Builder. It allows the instructors to form teams quickly and efficiently with an interventionist approach. The instructors can select from a well-defined set of rather general matching criteria, such as the timezone, a selection of tasks, or the participants' time commitment or preferred language.

sbw1 was accompanied by a survey and, additionally, user feedback was collected in the form of "I like... I wish..."[4] posts in the course discussion forum. Based on the results of the survey, the participants' feedback, and the analysis of the collected interaction data, we refined the initial prototype. For example, we added an introductory week to the following courses (*dt1*, *dt1-1*, *dt1-2*, *bmi1-1*), which was used to inform the participants about the team task, the collab spaces, and the peer assessment. In all of these courses, the teams were built at course start and worked in the same composition for the whole 5-6 weeks of the course. Each team was supported by a mentor. Therefore, the number of teams that have been admitted was limited by the availability of mentors.

In 2017, we delivered the first course that included an unmoderated, short (two-weeks), graded team task (*javaeinstieg2017*). At the end of the course we conducted a survey among the team-task participants. Furthermore, we interviewed 15 team-task participants for about one hour each. During some of these interviews, we asked the interviewee to demonstrate how they have solved certain tasks in their team. We observed

---

[4]I like, I wish is a simple feedback format that we use in our courses. The participants are encouraged to first list what they liked and then list what they think could be improved.

Figure 1. The average age (total course population) compared to the average age in teams. The size of the bubbles represents the relative size of each group. The maroon bubble for java1 represents ~20,000 participants.

these efforts to examine the workflows and detect usability problems on the platform. We then refined the *collab spaces* based on the results of the survey and the interviews. E.g. we added a video to explain the possibilities of the *collab spaces* and we renamed a set of menu items and removed features that haven't been used and rather confused the participants. We tested these improvements with a very small set of participants in an offline workshop, made some adjustments and ran the next large-scale experiments in the courses *bizmooc2018* and *java1*). We overhauled our surveys for each of the courses and conducted a new set of 15 one-hour interviews in the *java1* course (See [13] [14] [**?**] for previous results of our work). For the paper at hand, we have collected and analyzed the available participant data in all 15 courses that included a graded team assignment.

We have merged all available data sources of our platform—course reports, peer assessment reports, and team builder reports—to provide the most possible holistic view on the researched issue. From this raw data, we have generated three datasets:

1. A comparison between the participants who registered for the team tasks, those who did not register and the total course population, each aggregated on course level for all 15 examined courses.

2. A dataset aggregated on team level, for all 846 teams in the examined courses.

3. A complete dataset on user level for all 6246 team members.

In the course of our analysis, we have realized that some courses deviate substantially from our standard courses—e.g. the pilots and the workshops. The collected data from these courses, often rather distorts the image and leads to conclusions that are more likely the result of anything but the examined feature. We, therefore, have reduced the dataset and have filtered out pilots, workshops, and the course that has been offered to schools, in total we removed seven courses from our analysis. One commonality of the deselected courses is that they have been comparably small and hosted only a few teams. The dataset in total contains 846 teams. Our selection

of courses contains 703 teams (371 in the OOP category, 176 in the DT category, and 156 in the BI category).

For similar reasons, we have reduced the dataset even further in a few cases. Wherever it is relevant, we will explain this in more detail. If not stated otherwise, 703 teams have been examined.

We identified differences and commonalities between the total course population and the team-task participants. We analyzed the effects of certain team compositions on the teams' performance and dropout rates. Finally, we analyzed the effects of some platform modifications on team performance and team dropout rates.

To verify the validity of our conclusions, we have examined the group-wise distribution of the observations. In most cases we have stopped further investigations if not each of the groups have had a comparable size. Furthermore, we have double-checked each of the examined variables by comparing them separately on the course category or even course level. In many cases investigations of certain variables appeared to be promising in the beginning but then vanished into thin air.

TOTAL COURSE POPULATION VS. TEAMS
We compared the team members to the total course population to throw a light on the type of participants who register for the teams.

**Socio-Demographic and Geographical Background**
First we examined if the team members are representative for the total course population in terms of their socio-demographic and geographical background. The socio-demographic background data is collected in the user's profile. Providing this data is voluntary. About 35% of the team members and 25% of the course population have provided this data.

The geographical data is automatically collected based on the users' IP address whenever they access the course material. Therefore, we have 100% geographical data for the team members and about 60% geographic data for the total course population (this closely represents the overall show rate in the examined courses).

Figure 1 shows the average age of the team members versus the total course population. The maroon bubbles represent the total course population, the grey bubbles represent the team members. The size of the bubbles shows the relative size of the selection. The large maroon bubble at *java1* corresponds to ~20,000 participants. First of all, the graph shows that only a minority of the course participants registers for the team work. Second, the graph shows that there is no particular difference between the average age of the team members and the total course population. It has to be taken into account that in some of the examined courses, the age was a matching criterion. Therefore, the team members' age data is much more complete than the data of the other participants.

Figure 2 shows that the vast majority of the participants come with a bachelor's or master's degree[5]. This applies for both,

---
[5]To simplify things, we do not differentiate between a master's and it's older German counterparts Magister and *Diplom-Ingenieur*
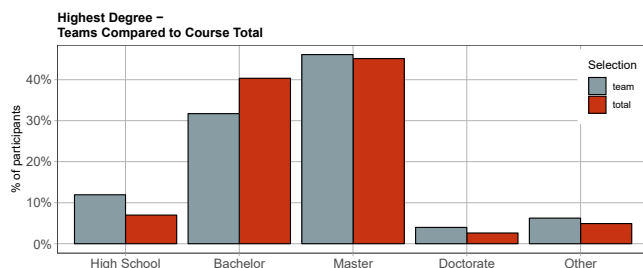
Figure 2. The majority of the participants in the examined courses has a Bachelor's or Master's Degree. These numbers are collected from the user profile data. To submit this data is voluntary. About 25% of the course population and about 35% of the team members have provided the data.

course total and team members. There are no significant differences between the courses. Asked about their career, 80% of the participants in the examined courses considered themselves to be professionals. About 10% are students, the rest are teachers, or academic researchers.

30% of the participants have the position of a technicians, closely followed by team leaders, project managers and department heads. About 60% of the participants have more than 10 years of professional experience, 20% have up to 10 years, another 20% have up to 5 years. An analysis of the data course by course showed that the results are very similar for all courses. Teams and course population also do not differ significantly in terms of first time platform users.

javaeinstieg2017 was offered in German language. This is mirrored by close to 100% participants from Germany in these courses. *bizmooc2018* was offered in cooperation with universities from Austria and Poland, which to some extent is reflected in the participants' origin. Except for *javaeinstieg2017*, all courses have particularly strong groups of participants from Germany, India and the United States of America. We have not found any fundamental differences between team members and total course population.

To sum it up, we can state that the socio-demographic and geographical background of the team members more or less parallels the background of the total course population.

Course Participation
Next to the socio-demographic and geographical background, we analyzed the course participation in terms of visited items, achieved points, active forum contribution, and course success in the form of certificates.

Visited Items
Figure 3 shows that team members in all examined courses, have visited a significantly higher percentage of items in each section than the total course population (We have observed the same phenomenon in the courses that have been removed from our selection). An item can be of type video, quiz, exercise, text, or assignment. The bubble size is defined by the standard deviation from the average value. Some of the sections have a very low percentage of item visits. E.g. Section 11 in *bizmooc2018* hosts a couple of video outtakes that have been
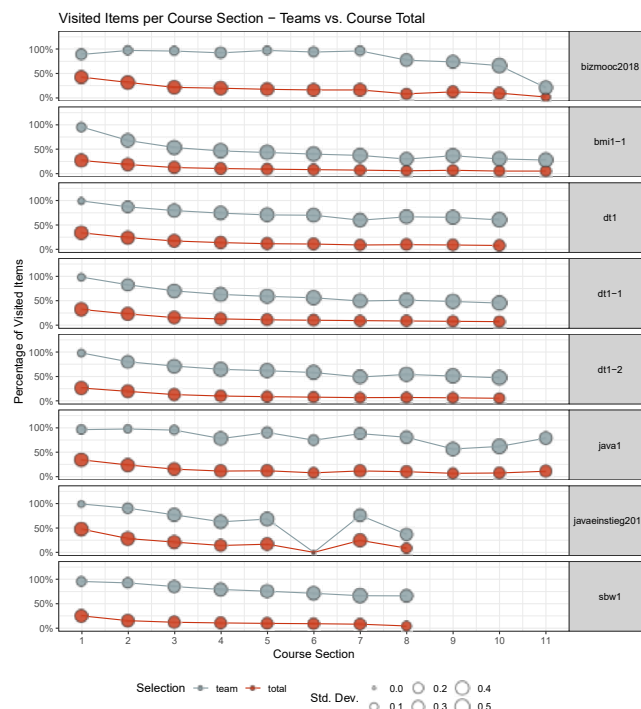
Figure 3. Team members have visited significantly more items throughout the course.

added after the end of the course. Sections 4, 6, 9, and 10 in *java1* have been optional and did not include an exam. In *javaeinstieg2017* Section 6 was an optional excursus.

Achieved Points and Course Success
Unsurprisingly, the achieved points in each section almost parallel the amount of visited items. More visited course items in combination with better results in exams and graded exercises result in higher course completion rates—measured in certificates (see Figure 4). In total, we can state with great confidence that it is generally the high performers who register for the team tasks.

Forum Activity
Finally, we examined the differences in the forum contribution between team members and the total course population. Figure 5 shows that team members are more active in the forums. To some extent this is expected, as the forum communication within the teams is included in this value. It is interesting that the forum contribution in the DT courses is particularly low among the team members. This is surprising as we would expect design thinkers to be a particularly communicative species. This might be an indicator that many of these teams have been local and were able to meet face to face—a quick-check of the Team Builder settings confirms that location has been a matching criterion, a quick check of the team data, however doesn't necessarily confirm this. Another possibility is that they have rather used the video chat than the forum for communication. To make a concluding statement, we will have to investigate this in more depth as we currently do not have data about the amount and length of the video chats.
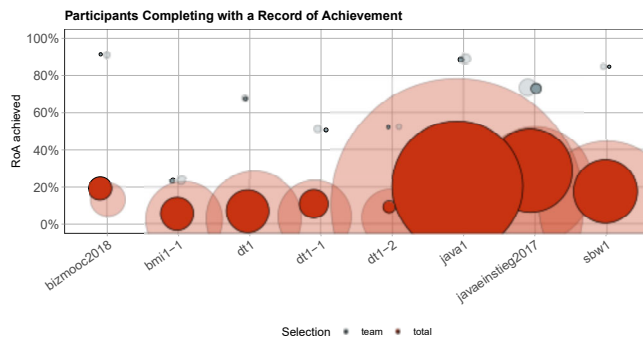
Figure 4. The y-axis represents the percentage of participants, who earned a certificate. The color of the bubbles represents the selection of examined participants (team members or course total). The size of the bubbles represents the size of the selection. The transparent bubbles represents the enrolled participants at the end of the course. The opaque bubbles represent the so-called "shows" at course middle. A "show" is an enrolled user, who has at least visited one course item. We, generally, measure our completion rates as the relation of certificates to shows at course middle. Example: The size of the transparent bubble at java1 represents ~20,000 enrolled participants. The opaque bubble shows that the course had a show-rate of about 50% and a total completion rate of slightly above 20%. The completion rate among the team members was close to 90%.)
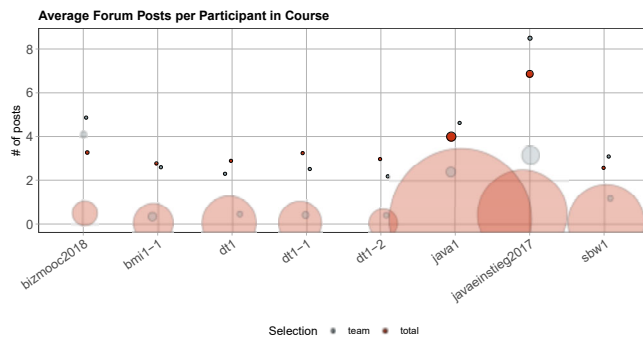

Figure 5. The transparent bubbles show the the average amount of forum posts per participant. The opaque bubbles show the amount of average forum posts per active forum contributor. Both separately for team members and course population. The size of the bubbles represents the relative size of each group. The transparent bubble at java1 represents ~20,000 participants.

ANALYSIS OF AGGREGATED TEAM DATA

Now that we have a basic understanding about the differences and commonalities between team members and the total course population, we can have a closer look at the aggregated team data and compare the teams' dropout rates and performance in regard to various aspects of the composition of their members. We have aggregated the data of all team members, to obtain a single observation for each team in all courses. As mentioned earlier, we then have removed some outliers from this list. Figure 6 shows an overview of the teams' success in in the examined courses. The color of the bar indicates in which phase of the peer assessment the majority of the team members has terminated to work on the task. The grey bar indicates that the majority of the team members has not reviewed the work of other teams—and most probably also has not contributed to the
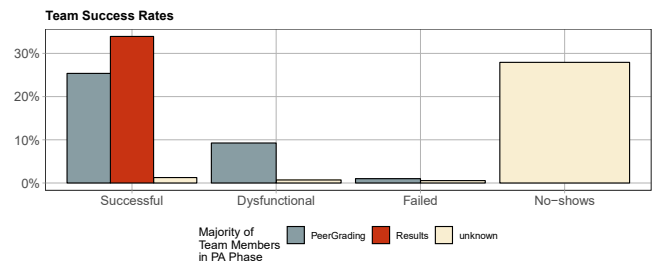

Figure 6. Team success rates. Successful teams: two or more members have finished the task, including the peer review. Dysfunctional teams: one team member has finished the task. Failed teams: The team has started to work on the task but none of the members have completely finished the task. No-shows: teams that have not started to work on the task.

team's submission—the maroon bar indicates that the majority of the team members have finished the task successfully in all phases. The beige bar indicates that the majority of the team has not started to work on the task at all[6]. We have defined a team as *successful* when at least two of the team members have received a grade for the task by submitting their work and reviewing the work of other teams. A team is *dysfunctional* if only one team member has finished the task. A team has *failed* if it has started to work on the task but none of the members has successfully finished the task. A team is considered a *no-show* if none of the members has started to work on the task.

Effects of Team Composition

We examined the teams' composition in relation to all socio-demographic aspects that are available in our data. In many cases, however, we have not found anything worth to report about. In some cases, our data turned out to be unfit to allow proper conclusions. For example, in some courses the team members expertise has not been collected at all, in some we had exclusively—or at least a vast majority of—teams with a heterogeneous mix of expertise. Results that at first sight might imply to be caused by a team's homogeneity or heterogeneity of expertise, did not withstand a closer examination. We, therefore, restrict ourselves here to some observations in the context of the teams' composition in gender, countries of origin, initial team size and the team members' commitment. Furthermore, we will examine the effects of mentoring and several platform modifications. For each of these aspects, we compared the teams' dropout rates, and the grade that they've received from their peers. We defined the dropout rate of a team as

$$dropoutRate = \frac{teamSizeStart - teamSizeEnd}{teamSizeStart}$$

The team size at the end is defined as the amount of team members that have reached the peer assessment's result phase. Except for the team dropouts, who do not receive any points,

---

[6]The few occurrences of teams that have finished the task with a majority of members that have not started the task, result from *javae-instieg2017*, where the instructors have removed inactive team members on request.
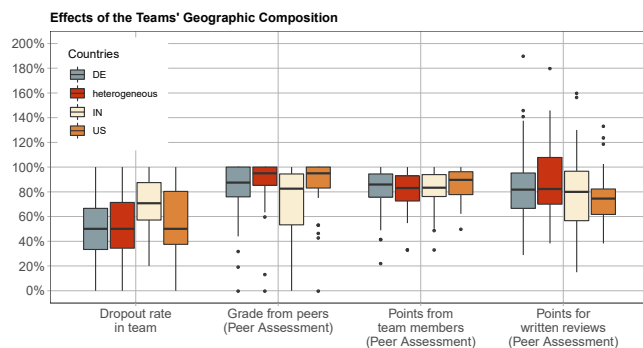
Figure 7. Team performance in terms of countries of origin. (451 teams)

Figure 8. Time commitment and team performance. (399 teams)

all team members receive the same grade for their team's submission. We have, therefore, aggregated the team grade as the maximum of the team members' grades. Additionally, we compared the points they received for their contribution from their team mates and the bonus points they received for valuable reviews[7]. Both are individual components and have, therefore, been aggregated as the mean of the team members' grades.

Socio-Demographic and Geographical

For the analysis of the teams' geographical composition, we have removed the data of *javaeinstieg2017* as this course was conducted in German language and, therefore, had a 100% German audience. Furthermore, we only examined the countries with the largest populations in the courses—Germany, India, and the US. To those we added the teams with a heterogeneous geographic background. We consider a team to have a distinct geographical background when the largest geographical group within the team comprises more than a third of the team members. Otherwise the team is considered to be geographically heterogeneous. All course categories are represented in this selection of courses. Each geographical group has a reasonably similar size within each category. We observed a significantly higher drop-out rate in teams with a majority of members from India, the performance of these teams also seems to be slightly lower (see Figure 7). Geographically heterogeneous teams on the contrary are performing very well.

Examining the teams' gender composition, we have observed a slight peak in the dropout rates of teams with a 80:20 male:female-ratio[8]. We, therefore, grouped the data by category to see if we can find differences. The peak shows significantly stronger in the Java courses and here, particularly, in *java1*. We will investigate this phenomenon in more detail in the future.

Initial Team Size

We grouped the initial team sizes into three categories: small (2-4 members), medium (5-7), and large (8-10). 81% of the teams were of medium size, 14% were large, and 5% small. The dropout rate in the small teams seemed to be significantly
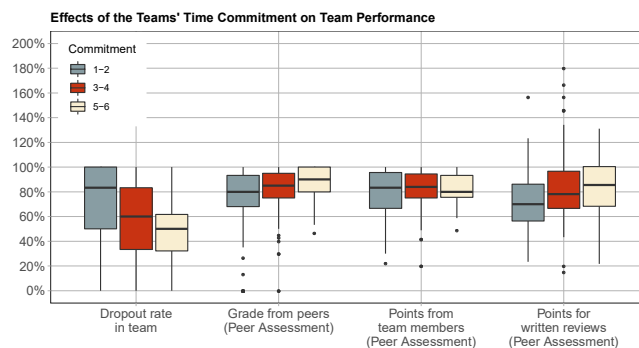
better than in the large or medium teams. However, the group-wise distribution of observations already indicates that this might be misleading. It also turned out that most of the small teams have been in one particular course. So, we had to let go of our initial conclusions. The same applies for the investigation of teams with even and odd numbers of team members.

Commitment

The participants' weekly time commitment for the given task early-on emerged as an important team matching criterion. In one of our previously published surveys ~40% of the participants selected a similar time commitment as the most important matching criterion [15]. It has been employed as the main matching criterion in the courses *javaeinstieg2017*, *bizmooc2018*, and *java1*. 61% of the examined teams committed to spent 1-2 hours per week on the given task, ~32% committed to 3-4 hours, ~7% committed to 5-6 hours. As we expected, the teams with lower time commitment have higher dropout rates and are less performant in terms of the grade received form their peers (see Figure 8).

Another indicator of a participant's commitment towards the course, is the amount of points that she has achieved in the exams and exercises before the team registration (PbT[9]). We have shown that participants who did not score 100% of the PbT will drop out of the team task with a close to 100% certainty. [14]

In *java1* we have used the Team Builder's filtering mechanism—based on the analysis of the *javaeinstieg2017* data (see [14])—to deny 65 of the 811 registered teamwork participants access to the team assignment. Additionally, we used the *PbT* as a matching criterion for some of the teams. Teams with a median *PbT* of all team members between 30% and 50% have been classified as *low*, teams with a median *PbT* of all team members between 50% and 80% have been classified as *medium* and teams with a *PbT* above 80% have been classified as *high*. Participants with a *PbT* of less than 30% have not been admitted for the team task. Furthermore, we defined the categories *homogeneous* and *heterogenous*. In homogeneous teams the difference between the lowest and the highest PBT is max. 10%. Figure 10 shows the distribution of

---

[7] As it is possible to write twice as much reviews as required, the last value can go up to 200%

[8] Given a team size of 5-6 members, this represents teams, where one woman is working with an otherwise all-male team
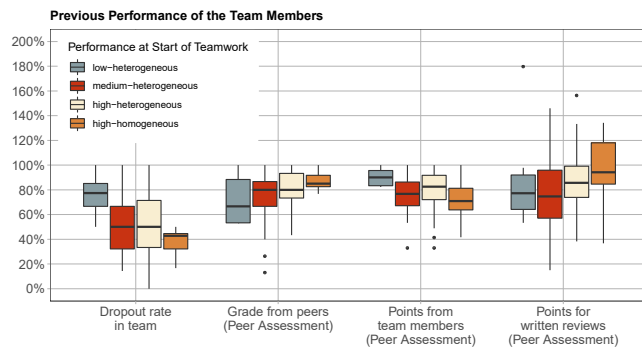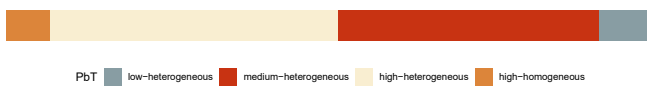
[9] Points by the time of team building.

Figure 9. Team performance in correlation to the results of the team members in previous course assignments and exercises (PbT) in course java1 (119 teams).

Figure 10. Distribution of the teams within the defined categories of previous results (PbT) in java1.

the teams in terms of these categories. The majority of teams have been heterogeneously-mixed high or medium performers. Additionally, we had a few heterogeneous-low-performers and a few homogeneous-high-performers among the teams. Figure 9 shows the performance results for these categories in comparison. It is no big surprise that the high performers have significantly less dropouts and have received the best grades from their peers. Adding the *PbT* not only as a filtering, but also as a matching mechanism, seems to be promising. Developing a proper recipe how exactly to match teams based on the *PbT* of their members, still has to be done.

Effects of Mentoring
Particularly in the courses with the long-running team tasks, the teams have been supported by mentors. Mostly, they have supported the teams organizationally. In a few cases they have also given feedback on the teams' intermediate submissions. The main disadvantage of mentors is that the concept does not scale and, therefore, the amount of teams has to be limited. Hence, we are interested in the effect of mentors on the teams' performance. For this purpose, we ran an experiment in *bizmooc2018*. In total 28 teams worked on this course's team assignment. We randomly selected five of them to be supported by active motivated human mentors. Another five teams were supported by a fake Robot[10]. Thirteen teams were supported by mentors that didn't have time to provide proper support[11]. The remaining five teams didn't receive any support at all. The fake Robot only sent regular announcements about upcoming deadlines to the team members and waited for their

---

[10]Fake Robot means that one of the authors pretended to be an automated tutor in the team forum.

[11]In some of the earlier courses some participants had complained about a similar situation, which is quite common among voluntary mentors.
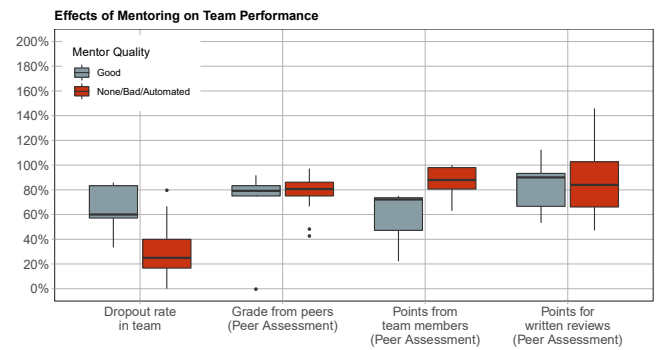
Figure 11. Is there a negative effect of mentoring on the teams' performance? (28 teams)

questions. The dialog was a one way street, however. None of the team members ever asked a question. The mentors who didn't have time for proper support, only sporadically sent deadline reminders. The *good* mentors tried to engage the teams:

Mentor 1 (Three teams):

In one of the teams, I started the first conversation by introducing myself and encouraging everybody else to engage in the conversation. One or two people only replied. I tried also to remind them with all the upcoming deadlines, ask questions about how they are proceeding but the response was extremely weak and it stopped after some time. To another team I was sending messages nonstop but none of the team members ever bothered even to say hello!

Mentor 2 (Two teams):

I served as a mentor for two of the BizMOOC teams. In this function, I welcomed them in the discussion forum of their team space, explained the available features and tools, and gave them some orientation about what they are expected to do. Furthermore, I presented myself and invited them to do the same.

Surprisingly, the teams of the *good* mentors performed significantly worse than the other teams. Figure 11 shows that the dropout rate in the mentored teams is about 30% higher than in the non-mentored teams. Based on this one experiment, it is hard to say if the well-intended efforts of our mentors scared-off the team members or if we have been unlucky with the random selection of our teams.

Effects of Platform Modifications
Interviewing the participants of *javaeinstieg2017*, and particularly observing their interactions with the platform's *collab spaces*, inspired many ideas to improve the toolset and the process. The most requested features by the participants have been a text chat and improvements in the platform's file-sharing abilities. So far, however, we haven't added any new functionalities.

- In contrary, we have removed features that turned out to be rarely used and confusing, such as *togetherjs*[12].

- We added a set of short videos to explain the features that support collaborative work and the mechanisms of the (team) peer assessment.

- We, generally, improved our communication strategy, and provided more detailed information at an earlier point of time.

- We restructured the *collab spaces'* navigation bar and re-named its items.

- We added an additional page to the video chat, to explain the participants that they need to schedule a meeting with their team mates first (and how to do that).

We measured the effect of these modifications by comparing the team performance results of *javaeinstieg2017* (pre-mod) with the results of *java1* (post-mod1) and *bizmooc2018* (post-mod2). We have selected this set of courses for the following reasons:

- *javaeinstieg2017* and *java1* are basically two iterations of the same course. *java1* has been offered in English language, *javaeinstieg2017* in German language. The teaching team has been the same in both courses. The size of the courses is also comparable.

- *javaeinstieg2017* and *bizmooc2018* had the same timing issue: the team task started more or less at the same day as the Easter holidays. In many teams half of the members were eager to start working on the task, while the other half was heading out for a vacation.

- The settings of these courses are comparable. The team task was optional in all courses. The time frame for the team task was similar. About two weeks to work on the task, plus one week to review and grade the peers.

In total, the selection consists of 399 teams. We would have preferred to add more pre-modification courses to the selection. However, all of them differ so much from the post-modification courses in duration and nature of the task, that we consider our choice to be the lesser evil. Instead, we have added the DT and BI courses to Table 1 as a reference. Table 1 shows that we have managed to more than double the amount of teams that passed. We have reduced the amount of no-shows significantly. The reference values from the DT and BI courses, however, show that other factors also have a strong influence. All of the DT courses and *bmi1-1* (BI) featured an introductory week to prepare the participants for the team task. Access to the team tasks was strongly limited due to the need for mentors. Only participants, who have solved an introductory exam have been eligible to apply for the task. The team task was an essential part of these courses, while in our selection it was only a bonus or add-on. All teams in the DT courses have been supported by professional or semi-professional mentors. Taken all this into account, we're positive that the platform modifications have to be considered as a success.

[12]https://togetherjs.com/

Courses	# Teams	Passed	Dysfunct.	Failed	No-show
Pre-mod	252	36.5%	15.1%	3.2%	45.2%
Post-mod1	119	78.2%	10.1%	0%	11.8%
Post-mod2	28	78.6%	14.3%	0%	7.1%
BI	128	55.5%	7.8%	0.8%	35.9%
DT	176	83.5%	3.4%	1.1%	11.4%

Table 1. Pre-mod (javaeinstieg2017) and post-mod (java1 and biz-mooc2018) courses. Business innovation (BI) and design thinking (DT) courses have been added as a reference.
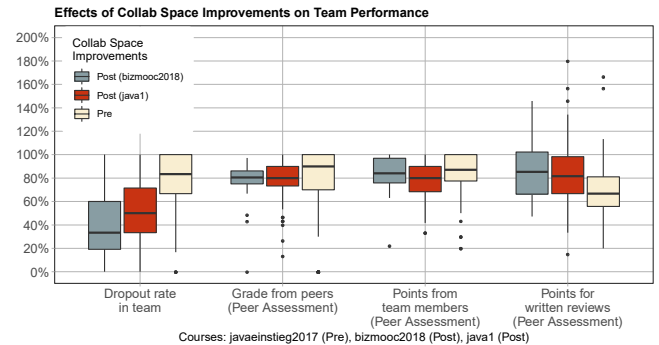
Figure 12. Improved team performance in java1 and bizmooc2018 (post-modification) compared to javaeinstieg2017 (pre-modification) (399 teams)

Figure 12 compares the dropout rates and team performance within our selection. The dropout rate in the teams has plunged from more than 80% in *javaeinstieg2017* to about 50% in *java1* and about 30% in *bizmooc2018*. Table 1 might imply that this is mainly due to the decreasing amount of no-shows. We, therefore, compared the start and end sizes of those teams that have started to work on the task in the pre- and post-modification courses. While the teams in all courses started with an average size of about six members, the average team size at the end of the task was 2.2 members in the pre-mod course vs. ~3.5 members in the post-mod courses. The team performance in terms of the grades that the teams have received from their peers, has slightly decreased. If this was caused by an actual lower performance of the teams in these courses or by a more strict grading of the peers will have to be examined in more detail. The better results for the written reviews in the post-mod courses, are at least partially caused by an increased number of ratings for the reviews[13]. While in the pre-mod course only 41%[14] of the participants have received a rating for their review, in the post-mod courses 53-63% of the participants have received a rating.

CONCLUSION AND FUTURE WORK

The paper at hand is part of a long term study that aims to understand and improve scalable graded team-based assignments in MOOCs. A particular challenge in this context is the loose coupling of the participant to the course as well as

[13]Rating a review is optional. Encouraging the participants to make use of this option is one of our goals.

[14]41-46% if we also include the business innovation and design thinking courses

to the providing institution. This loose coupling prohibits approaches to form perfect teams, such as e.g. questionnaires in the form of Belbin tests, and often results in high dropout rates within a course. While, in general, we do not consider high dropout rates in MOOCs to be a big deal, they do constitute a major challenge in the context of team work. To predict and prevent dropouts, we need a solid understanding of our participants. We, therefore, have compared the team task participants to the general course population and analyzed the aggregated team performance data in selected courses. We have shown that while the socio-demographic and geographical background of the team task participants more or less mirrors the total course population, mostly high performing participants are registering for the team tasks. We have also shown that lower performing participants are very likely to dropout from the teams and, therefore, recommend to establish an entry test for the team task before the team building process starts. Finally, we have shown that a successful team assignment depends on getting the details right. An intensive communication with—and observation of—our participants enabled us to successfully improve the platform's collaboration features and the communication of the processes. For future work, we have identified several topics that appear to be promising for digging deeper. For example, it could be interesting to have a closer look at teams where one woman is confronted with an otherwise all-male team. Employing previous course results not only to filter participants, but also as a feature to match teams seems to be another approach that is worth some additional effort.

REFERENCES

1. Anne. 2014. Curso Evaluación de Decisiones Estratégicas: Vive el desafío MOOC en español. https://novoed.com/blog/2342/curso-evaluacion-de-decisiones-estrategicas-vive-el-desafio-mooc-en-espanol/. (2014). [Online; accessed 10-01-2018].

2. R. Meredith Belbin. 2012. *Team Roles at Work* (2 ed.). Routledge.

3. T. Brown. 2009. *Change by Design: How Design Thinking Transforms Organizations and Inspires Innovation*. HarperCollins.

4. Seunghoo Chung, Jr. Robert B. Lount, Hee Man Park, and Ernest S. Park. 2018. Friends With Performance Benefits: A Meta-Analysis on the Relationship Between Friendship and Group Performance. *Personality and Social Psychology Bulletin* 44, 1 (2018), 63–79.

5. Susan Brown Feichtner and Elaine Actis Davis. 1984. Why Some Groups Fail: a Survey of Students' Experiences with Learning Groups. *Organizational Behavior Teaching Review* 9, 4 (1984), 58–73.

6. A. Ju, X. Fu, J. Zeitsoff, A. Hemani, Y. Dimitriadis, and A. Fox. 2018. Scalable Team-Based Software Engineering Education via Automated Systems. In *2018 Learning With MOOCS (LWMOOCS)*. 144–146.

7. René F. Kizilcec. 2013. Collaborative Learning in Geographically Distributed and In-person Groups. In *Proceedings of the Workshops at the 16th International Conference on Artificial Intelligence in Education AIED 2013, Memphis, USA, July 9-13, 2013*.

8. Roger L Martin. 2009. *The design of business: Why design thinking is the next competitive advantage*. Harvard Business Press.

9. Jonathan McKinsey. 2015. *Remote Pair Programming in a Visual Programming Language*. Technical Report UCB/EECS-2015-139. University of California Berkeley, Berkeley.

10. Marguerite McNeal. 2016. Online Classes Get a Missing Piece: Teamwork. https://www.edsurge.com/news/2016-09-28-online-classes-get-a-missing-piece-teamwork. (2016). [Online; accessed 14-10-2016].

11. Carolyn Penstein Rosé, Oliver Ferschke, Gaurav Tomar, Diyi Yang, Iris Howley, Vincent Aleven, George Siemens, Matthew Crosslin, Dragan Gasevic, and Ryan Baker. 2015. Challenges and Opportunities of Dual-Layer MOOCs: Reflections from an edX Deployment Study. In *Proceedings of the 11th International Conference on Computer Supported Collaborative Learning*. 848–851.

12. Junko Shimazoe and Howard Aldrich. 2010. Group Work Can Be Gratifying: Understanding & Overcoming Resistance to Cooperative Learning. *College Teaching* 58, 2 (2010), 52–57.

13. Thomas Staubitz and Christoph Meinel. 2017. Collaboration and Teamwork on a MOOC Platform: A Toolset. In *Proceedings of the Fourth (2017) ACM Conference on Learning @ Scale (L@S '17)*. ACM, New York, NY, USA, 165–168.

14. Thomas Staubitz and Christoph Meinel. 2018. Team Based Assignments in MOOCs: Results and Observations. In *Proceedings of the Fifth Annual ACM Conference on Learning at Scale (L@S '18)*. ACM, New York, NY, USA, Article 47, 4 pages.

15. T. Staubitz, H. Traifeh, and C. Meinel. 2018. Team-Based Assignments in MOOCs - User Feedback. In *2018 Learning With MOOCS (LWMOOCS)*. 39–42.

16. Miaomiao Wen. 2016. *Investigating Virtual Teams in Massive Open Online Courses: Deliberation-based Virtual Team Formation, Discussion Mining and Support*. Ph.D. Dissertation. Carnegie Mellon University.

17. Z. Zheng, T. Vogelsang, and N. Pinkwart. 2015. The Impact of Small Learning Group Composition on Student Engagement and Success in a MOOC. In *Proceedings of the 8th International Conference on Educational Data Mining (EDM 2015)*. Madrid, Spain, 500–503.